

The NPT and its black sheep

The turmoil in what used to be the Soviet Union has distracted attention from projects for arms control, but the opposite should be the case.

Mr James Baker, the US Secretary of State, did not hide his alarm at North Korea's nuclear ambitions on his swing (instead of President George Bush) the other week through Tokyo, Seoul and Beijing. Meanwhile, US sources are busily quoting what is called intelligence information in their contention that North Korea is well on the way towards the manufacture of plutonium from reactor fuel, and that its first nuclear weapons may be only months or a few years away. He was right to say that such a development could be a calamity for the Pacific region. Not since the United States itself considered using nuclear weapons in the Korean War has there been such a serious threat to the fragile stability of the Korean peninsula and its neighbours (which include Japan). But that, sadly, is only part of an alarming story.

Nuclear weapons abound, but so now do putative nuclear weapons states. That is one consequence of the revelations in the past few months about Iraq's nuclear programme. Although it is not inherently surprising that a government well supplied with cash and with the means to be single-minded should have gone so far in the replication of a technology now half a century old, Iraq's doings have inevitably converted every non-signatory of the Nuclear Non-proliferation Treaty (NPT) into a putative nuclear power. And the list of them is too long for comfort. Apart from North Korea, there are Israel and several of its neighbours (including Egypt and Saudi Arabia), India and Pakistan, Argentina and Brazil. (The good news is merely that South Africa is now a signatory.) Although many of these states may have set aside their nuclear ambitions (like Argentina and Brazil, for example), they now have only themselves to blame if their neighbours behave as if they were nuclear weapons states.

There is worse than that. The bilateral agreements between the United States and the Soviet Union would be a greater comfort if it were clear which surviving elements of the Soviet government will have continuing responsibility for them. (Apparently there are also technical problems, such as the difficulty of dismantling weapons whose designs have been forgotten or mislaid.) As things are, there are no formal arrangements that constrain the British and French nuclear forces, which are said to have a strategic role. And China, which exploded its first bomb a quarter of a century ago and which is probably by now a major player in the field, is similarly unconstrained.

Baker, evidently hoping to win from Beijing some help with North Korea, has good reason to be disappointed with China's refusal to see a ganging-up against North Korea.

There is no easy solution. The view (see *Nature* 353, 483; 1991) that the United Nations could and should compel putative nuclear powers to sign the NPT is likely to be frustrated by the protection offered by major powers to their client states. (Would the United States be more or less likely to veto such a resolution in regard to Israel than China in regard to North Korea?) Middling powers such as Britain and France, which are permanent members of the UN Security Council, would similarly use their vetoes to fend off coercion of themselves. The outstanding difficulty is that, while China remains entirely immune from dialogue on strategic arms that, during the 1970s, drew the old Soviet Union into arms control, massive strategic forces will persist.

Yet the NPT will lapse four years from now unless its signatories persuade themselves to soldier on. It is ironical that while the major powers (the United States and the Soviet Union) have now done what the NPT requires of them (to negotiate "in good faith" strategic arms control), many of the signatories are more likely now to be given pause by the emergence of nuclear power among countries like themselves. The review conference in 1995 is unlikely to be less rowdy than its predecessors. That prospect should persuade even the major powers to see sense. Let us hope they do so in time to save the NPT. □

Hidden buildings

Overhead charges on research grants subsidize new buildings. The practice needs to be examined.

THE National Institutes of Health (NIH) have done the sensible thing in seeking to simplify accounting for indirect costs (see story on page 258), but that will not deal with one important issue — how new construction is to be financed on university campuses. Until about 15 years ago, in the United States, when the federal government was building the world's finest biomedical research enterprise through the National Institutes of Health (NIH), Congress readily provided funds for bricks and mortar to house research. Then, as the steep climb in NIH funds



began to level off, money for construction all but disappeared from congressional appropriations, the outrage of medical school administrators notwithstanding. Yet new construction continues, if at a slower pace. Where does the money come from?

From indirect costs, of course. Universities are allowed under the rules to charge to research grants a portion of the costs of constructing and financing new buildings. What this means is that institutions confident that their researchers would be well supported have been able to build buildings speculatively, knowing that the costs would eventually be requited by the indirect costs with which research grants are loaded. The results have often been dramatic. One research hospital, for example, used to charge a 30 per cent overhead rate until it built a new building, when overhead charges (on every grant, whether or not the new space was used) jumped to 70 per cent.

The system, entirely legal and above board, is now being scrutinized closely, in the wake of the more flamboyant scandals that have come to light. Several other abuses have been uncovered in the allocation of what is called the administrative overhead, which has been rising steadily over the past decade. The use of the indirect cost pool for the amortization of buildings seems to be one of the chief forces driving up the overhead rates, at some institutions now 80 per cent of the value of research grants. Yet nobody pretends that new buildings are not needed. Too many research buildings in the United States are now 40 to 50 years old, in desperate need of rejuvenation or replacement. It may be possible to start computer companies in a garage, but garages are not the place for contemporary science.

What is to be done? Congress must first understand that its pretence that it supports research, but does not build buildings, is a misconception. Having willed the projects, does it not have a responsibility for the settings in which they will be carried through? To be sure, there are advantages in insulating decisions on buildings from the geographically sectional interests rife in Congress, but none in a system that makes research seem more costly than it is and which rewards the institutions whose speculative builders are luckiest. What the US research enterprise needs is analysis of the need for research facilities in the twenty-first century and a commitment from the government to provide them. □

Death with dignity

A proposal in Washington state to legalize euthanasia has begun an important debate even though it did not become law.

THE idea that the terminally ill should be allowed to "die with dignity" has become not only a moral but a potent political issue, first in the Netherlands, now in the United States, where the sometimes mindless use of high technology to sustain otherwise lifeless bodies is all too common.

That explains the voter initiative on the ballot in the state of Washington earlier this month that would have legalized euthanasia. The proposition was narrowly defeated. If it had become law, Washington would have been the first jurisdiction in the Western world to sanction physician-assisted death. Even the Dutch, long tolerant of help in dying, have not explicitly legalized euthanasia. And the failure of the Washington initiative in no way settles the matter, but merely puts it off until the next statewide election.

Even those who accept that physicians should, in certain circumstances, accede to the wishes of a terminally ill patient that he or she should die agree that the issues are complex and that the slippery-slope argument against euthanasia has force. What separates euthanasia for the truly terminally ill and mercy killing of the hopelessly sick (Alzheimer's patients, for instance)? Washington's draft law contained several prohibitions of the slippery slope. First, it would have required a determination, to the extent that medicine can judge, that the patient had less than six months to live. Second, it would have applied only to mentally competent patients, those alert enough to ask for aid in dying. Strictly kept, these provisions would have made for sound protection.

So why did the proposition not succeed? One influence, during the Washington ballot, seems to have been a last-minute television advertisement featuring a cancer patient who claimed that, had physician-assisted dying been legal when her disease was diagnosed seven years previously, she might have chosen to die. That claim, of course, overlooks the need specified by the draft legislation that a determination of imminent death should have been reached objectively. Yet the advertisement had obvious emotional appeal; it may have influenced voters whom pollsters had predicted were planning to vote YES.

The important and neglected question, which points to the reason why this debate will not go away, is why so many people are leaning towards legalized euthanasia, even in the face of religious scruples. Why, for that matter, has the Hemlock Society's book *Final Exit* become a bestseller in the United States? The answer is that modern medicine, with its devotion to respirators and other forms of high technology, has lost its bearings.

Most terminally ill people do not rush headlong into suicide. What makes it an attractive option is a deeply held fear that, when death is at hand, eager doctors armed with feeding tubes and breathing machines will fend it off — painfully, at great cost, but only for a while. It may be relevant that, in Canada and Europe, the high technology is less conspicuous, as is the likelihood that a family will bring a malpractice suit if doctors fail to use the very technology that so many patients fear. The physicians' common objection to euthanasia on the grounds that the Hippocratic oath compels them to "do no harm" overlooks the reality that, often, patients perceive modern medical technology as harm in itself. At least until the next election, Washington can be pleased to have made these issues respectable in polite conversation. □

Cholera epidemic traced to risk miscalculation

- Cancer fear led to halting of chlorination
- Uncertainties in balancing risks

Washington

A DECISION by Peruvian officials not to chlorinate much of the country's drinking water, which was based on studies by the US Environmental Protection Agency (EPA) showing that the chlorine may create a slight cancer risk, is being blamed for the devastating cholera epidemic that is now sweeping Peru and a dozen other countries in South and Central America.

Since the first incidents of cholera were identified in January, more than 300,000 new cases have been reported, mostly in Peru. Statistics released earlier this month by the Pan American Health Organization (PAHO) show that the epidemic has claimed 3,516 lives. PAHO officials believe that the bacteria first arrived with a Chinese freighter, which released its apparently contaminated bilge water into the harbour at Lima, Peru. The bacteria quickly made their way to the shellfish and fish, probably reaching humans first in the form of *ceviche*, a raw seafood dish popular in Peru. But once the disease appeared in humans, it quickly moved into the water supply, infecting many times as many people as might have otherwise been exposed by person-to-person contact.

US and international health officials last week blamed Peruvian water officials for a gross miscalculation in not chlorinating the entire water supply. Although Peru has good water-filtration technology and pumps safe water into the drinking water system, old pipes and open unchlorinated wells appear to have allowed the cholera bacteria to enter the water supply after filtration. Chlorine is a disinfectant and could have protected the water supply even if it were exposed to bacteria.

Yet chlorine can also react with by-products of organic decay in the water to create several suspected carcinogens, including chloroform. A class of these chlorine-based compounds, called trihalomethanes (THMs), are currently regulated by EPA, which requires that their concentration in major water systems be less than one part per billion (p.p.b.). EPA studies in the 1970s found that a 100-p.p.b. level poses a cancer risk of about 1 in 10,000.

Since then, EPA has wrestled with the question of how to balance the cancer risk of chlorination with the microbial risk of no disinfection at all (chlorine alternatives, such as ozone gas, are expensive and may have even more serious health effects). Current US regulations set a 100-p.p.b. limit for THM chemicals, although one study suggests that even that level may cause 700 extra cases of cancer each year in the United States. Yet most epidemiologists agree that a relatively small risk of cancer is preferable to the possibility of a microbial epidemic.

During the 1980s local water officials, citing the EPA studies of chlorine's cancer potential, decided to stop chlorinating many of Lima's wells. This has now raised serious questions about both EPA's risk assessment and the way it has been communicated to the rest of the world. Given the uncertainties of risk assessment and the difficulties in balancing microbial and cancer risks, researchers ask whether EPA should have given more emphasis to the disaster potential of not disinfecting water supplies.

Even if EPA was sending mixed signals, Peru appears to have heard only what it wanted to hear. Frederic Reiff, PAHO's

regional director for water quality, says the decisions may have been based more on the practical and economic difficulties of chlorination than on analysis of the risks. The EPA studies "were one of a number of excuses they used to not chlorinate their water," he says. Robert Clark, director of the EPA Drinking Water Research Division, adds, "They knew the dangers. I think that they were simply using the EPA position, so they could turn around and point the finger at us and say, 'Well, they told us not to.'"

This sobering case of risk assessment gone wrong is forcing US and international health officials to come to grips with the flaws in what most agree is a haphazard process of balancing real and theoretical public health risks. At a risk assessment meeting at the National Academy of Sciences in Washington last week, researchers and health organizations urged US officials to reexamine their analysis of chlorine's cancer risk in the light of the South American epidemic.

"Chlorination and disinfection of the water supplies are the public health success story of the century," said Carol Henry, director of the International Life Sciences Institute's (ILSI) Risk Science Institute. "To start altering this in some way has very grave and immediate consequences. I don't think we've looked at this with any rational or reasoned approach."

Next August, ILSI will convene an international conference to discuss the problems of balancing chemical and microbial risks in water supplies. Meanwhile, the EPA, which had promised to update its chlorination standards by this year, is still wrestling with the basics: in the absence of a viable alternative to chlorination, the agency is reluctant to reduce the amount of permissible chlorine levels and risk a Peru-like epidemic. EPA officials say that they need more research to estimate chlorine's real-world risk. New regulations are not expected before the middle of the decade.

Christopher Anderson

Splitting the difference on risk

FOR more than ten years, two US government agencies — both charged with protecting public health, and both trying to calculate human risks based on animal tests — have disagreed on one of the most basic elements of risk assessment. The Food and Drug Administration (FDA) converts rat risks to human risks by multiplying by the difference in weight between the two species. But the Environmental Protection Agency (EPA) uses the difference in body surface area. For the same chemical, FDA would get a human risk estimate 13 times the EPA calculation. Which agency was right? No

one cared. "It was a matter of turf," says EPA deputy administrator Henry Habicht.

Next month, that situation will end. Conceding that US risk assessment policy is a mess, a panel of high-ranking government officials have embarked on an ambitious programme of reforms. And their first victory is the 'scaling problem'. By the end of the year, both EPA and FDA are expected to publish new procedures setting a common rule for scaling from animals to people: multiply by the difference in body weight to the three-quarters power.

Last week, the government risk assessment panel — part of the Federal

Coordinating Committee on Science Engineering and Technology (FCCSET) — held its first public meeting. More than 40 commentators — from chemical industry representatives to environmental activists — called for, among other things, better research on microbial risks (see story this page), neurotoxicity, exposure durations, animal models (fish were recommended as a replacement for rats) and risk communication.

Nearly all the commentators asked for more data and less politics, something that the FCCSET panel promised to keep in mind. **C.A.**

Deciding on indecision

London

LAST week's gathering of European ministers in Munich may have failed in its task of deciding the future of the European Space Agency (ESA) in manned space-flight, but most delegates went home satisfied that they at least avoided a catastrophic disintegration of ESA's two flagship projects — the reusable spaceplane Hermes, and Columbus, Europe's contribution to the US-led space station Freedom. For the time being, the ministers decided, both will continue, although with slightly reduced budgets.

The cost of saving the projects, however, is that Europe's ambition to become an independent power in manned space-flight has taken a body blow: if ESA officials cannot pull in extra funds by persuading the Soviet Union and Japan to collaborate in the agency's manned programme, the German government may yet act on its threat to withdraw from Hermes when the ministers meet in Spain a year from now — an action that would kill the project.

ESA director-general Jean-Marie Luton was seeking backing for a plan to complete Hermes for \$7,600 million (at 1990 prices) and Columbus for \$5,300 million. Although Luton did not receive this endorsement, ESA's managers are relieved at the outcome of last week's meeting. Not only were the reluctant Germans kept in the Hermes project, despite a 40 per cent cost overrun since the last ESA ministerial meeting in 1987, but a threatened revolt of the smaller ESA nations was also averted. Norway, which was expected to lead this protest by withdrawing its 0.2 per cent contribution to the Hermes budget, was placated by assurances that its aerospace industry will receive a fair share of the contracts to build the spaceplane.

The central question was whether Europe's grand ambition to develop an autonomous manned space programme is still realistic. This idea has long been championed by France, and in the relative economic plenty of 1987, the other large ESA states — apart from Britain — were prepared to agree. But in today's harsh economic climate, enthusiasm for expensive manned space projects is waning.

Germany, the second biggest contributor to ESA after France, has reconsidered its commitment to big space programmes because of the budget squeeze forced upon the German government by the need to invest in the former East Germany. Research and technology minister Heinz Riesenhuber had a clear brief in Munich not to agree to Luton's extended budget

for Hermes. With Riesenhuber threatening publicly to pull out of the project if France, Hermes' main supporter, forced the issue, last week's compromise was perhaps the best ESA could hope for.

The obvious route for savings on Hermes is to interest the Japanese and the Soviets in taking a share in the project, but this will take time to arrange. The alternative to the present year of limbo could have been the loss of both Hermes and Columbus. Although Columbus is still favoured by Germany, a German withdrawal from Hermes might have led to the French making retaliatory attacks against the Columbus budget.

Under the Munich agreement, Hermes and Columbus will continue to inch forwards, but next year's budget for each must be trimmed by some \$40 million — more than 10 per cent below the figure Luton had hoped to spend. ESA officials will be able to award some contracts to build hardware for the projects, but aerospace companies accepting them have no guarantee that these will last beyond the end of 1992. In total, Luton will have to cut 5 per cent from ESA's 1992 budget, although spending on its space science programme, into which all the ESA members are obliged to pay, will be protected from any cuts.

Meanwhile, Germany's independent space science programme, which had been threatened with cuts if Riesenhuber failed to control German spending on ESA projects (see *Nature* 354, 97; 14 November 1991), should be secure for at least another year — although Germany has not yet decided whether it can afford to contribute in 1992 to the CRAF mission, a joint project with the United States to study a comet and the asteroid belt.

Despite having forced the French to compromise, the German research ministry is still worried that its demands for

savings will not have been met by the time ministers gather in Spain next year. Gottfried Greger, a space policy official at the research ministry, says ESA must begin talking with the Soviets and the Japanese immediately to determine what each can offer ESA. But he fears that ESA's management is reluctant to relinquish its dream of independence in manned space-flight. "We are a little afraid that ESA doesn't want to understand our message," says Greger.

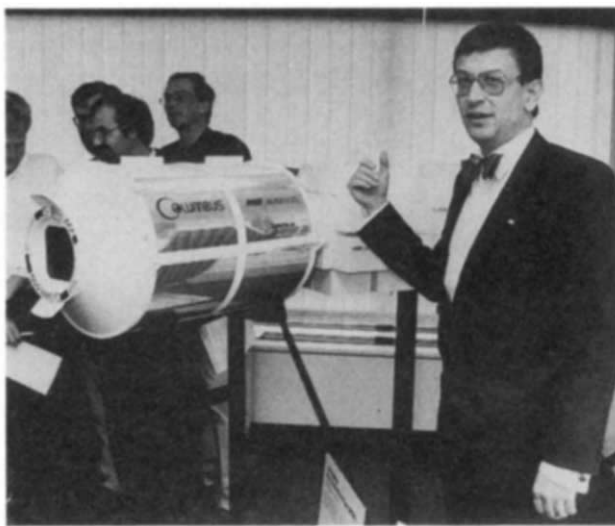
The Soviet space programme, with considerable experience in manned space stations through its Mir programme, may be able to contribute to Columbus. And Japan is planning its own (unmanned) spaceplane, called Hope (see *Nature* 351, 680; 27 June 1991), and so may have technology useful for Hermes. But precisely how new partners can become involved is still uncertain, and, in the case of the rapidly unravelling Soviet Union, ESA may also have problems finding the correct officials with whom to negotiate.

One thing, however, is clear after last week's meeting. The ESA member states are unlikely to allow ESA's programme to run for four years at a time without scrutiny by their ministers, as they have in the past. "We think it might be useful to have more regular [ministerial] meetings," says Greger, provided that these are low-key affairs. Greger points to the annual gatherings of European research ministers to discuss the Eureka applied research and development programme, which keep Eureka in check without causing the public controversy that preceded last week's ESA conference.

But would not regular ministerial scrutiny of ESA's programmes introduce the same uncertainty that afflicts the US National Aeronautical and Space Administration's long-term planning? Only this year, ESA director-general Luton appeared before a House of Representatives committee complaining bitterly about the possibility that US funding for the

space station Freedom could be cancelled, as part of the annual Washington budget round (see *Nature* 351, 428; 6 June 1991). Greger agrees that a long-term ESA strategy is essential, but says that this must be balanced against the need to match spending to shorter-term economic realities. Achieving this balance will be ESA's most difficult challenge over the next few years, if Columbus and Hermes survive beyond next year's Spanish showdown. In the long run, the losers from the Munich meeting are likely to be Europe's aerospace companies, who may find that guaranteed multi-year contracts to build hardware for ESA are a thing of the past.

Peter Aldhous



Riesenhuber retreats on big space programmes.

JESSI

Losing the 'R' from R&D

London

THE Joint European Submicron Silicon Initiative (JESSI), the seven-year research and development programme designed to counter the growing dominance of the Japanese electronics industry, is being shorn of almost all of its basic, long-term research. With JESSI forced to trim its planned budget for 1992 by 25 per cent, researchers working on basic semiconductor research — most of whom are based in universities or other nonprofit institutes — will bear the brunt of the cuts.

Two years after its inception, JESSI has run into financial difficulties. According to JESSI officials, the main problem is the failure of the European Commission to provide more than a third of its promised contribution (see *Nature* 354, 5; 7 November 1991). When JESSI was planned, one quarter of its \$5,000-million budget was supposed to come from the Commission, a similar sum from the individual European governments interested in the programme (Belgium, Britain, France, Germany, Italy and the Netherlands have since signed up), and the remaining half would be supplied by the European semiconductor industry itself.

For an expensive research and development programme to be forced to cut back is not unusual. But JESSI's budget shortfall will alter the balance of the entire project. Planned as a multilayered scheme to address most aspects of semiconductor research and development — from basic research not expected to yield products for a decade or more, to improvements in current industrial semiconductor manufacturing processes — JESSI may now become purely a product development programme, concentrating on technologies and applications for semiconductors that are likely to come to the market within the next five years.

The problems for JESSI's basic research programme began a year ago, when the German government, as part of its general retreat from international research collaboration in the face of the cost of German unification, decided it could not pay for JESSI's basic semiconductor research beyond the end of 1991. The other five countries quickly followed suit, arguing that the national governments had to take a consistent line, leaving public funding for the programme entirely in the hands of the European Commission. When the problems with the Commission's contribution became apparent earlier this year, JESSI's governing board realized the basic research programme was going "to run straight into a brick wall", says Ron Lawes, from the UK Science and Engineering Council's Rutherford Appleton Laboratory, who serves on the JESSI board.

Lawes, who has responsibility for

JESSI's basic research (which was planned to consume some 15 per cent of the total JESSI budget), may soon find that this role evaporates to nothing. Even the jewel in the crown of the basic research programme — a project to develop microchips whose smallest features are only 0.25 micrometres wide — has no firm guarantee of funding. (This project is more ambitious than the efforts of the Sematech consortium, JESSI's counterpart in the United States, which is aiming at 0.35-micrometre devices). Almost one-half of the 0.25-micrometre project has already been lost, and if the project is to be funded at all, it must compete with many hundreds of other proposals chasing after grants from the European Commission's heavily oversubscribed ESPRIT information technology programme. "We're optimistic that we'll get an answer by Christmas," Lawes says. But with the Commission keeping JESSI in the dark about its deliberations, he is less optimistic about what that answer might be. Other projects have been cast aside entirely, including research into semiconductor 'packaging' — how to integrate new and faster microchips into working electronic devices.

The collapse of JESSI's basic research programme is expected to reduce drastically the semiconductor research effort in Europe's universities and non-commercial research institutes. Unlike most areas of JESSI, where the majority of the work is planned for industrial laboratories, about three-quarters of the scientists and engineers who expected to be employed on JESSI's basic research are based outside of industry.

But JESSI is an industry-led programme, designed to strengthen the European semiconductor industry, rather than to subsidize academic researchers. "It's fair to say that Europe's problems in the industry are not basic research problems," says Michael Borrus, from the Berkeley Roundtable on the International Economy, who follows the electronics industry. So how will the loss of most of JESSI's basic research affect the programme's main mission of responding to the success of the Japanese industry?

As might be expected, the line from industry is slightly different from that offered by Lawes and Radelaar. "If we don't survive the next two years, I don't care about the next ten years," says Hartwig Bierhenke, a research and development planner with the German electronics company Siemens, one of the biggest industrial players within JESSI. "If there is a lack of money, we have to concentrate on the nearer goals." Significantly, Bierhenke sits alongside Lawes on the management board responsible for JESSI's basic research programme.

Nevertheless, the European semiconductor industry's preoccupation with its short-term goals contrasts strongly with the long-term vision of the Japanese industry. And the loss of JESSI's basic research will have one tangible effect. The basic research programme was designed in part to provide technologies that could be developed into products under a successor to JESSI. If JESSI sheds its basic research, then a follow-up programme, after JESSI expires in 1996, looks extremely unlikely.

Peter Aldhous

GENETIC ENGINEERING

Rifkin wins suit

Washington

THE Foundation on Economic Trends and its president, Jeremy Rifkin, last week won a court victory over the Department of Defence (DoD) which will force the military to prepare studies on the environmental effects of their research programme to find defences against biological weapons.

The US District Court for the District of Columbia ruled that DoD must file an environmental impact statement (EIS) on its Biological Aerosol Test Facility at the Dugway proving grounds within six months. The court also ruled that DoD must prepare environmental assessments at five other facilities where the Army or DoD contractors are conducting research on biological toxins in the two highest risk categories, known as BL-3 and BL-4.

DoD spokesman Rick Thomas says the Army has already prepared an EIS for Dugway and is in the process of finishing environmental assessments at the other facilities. "Nothing in this agreement forces the Army to do anything we weren't planning to do in the first place," he says.

Officials at the National Institutes of Health (NIH), which conducts BL-3 and, occasionally, BL-4 experiments, are concerned that the ruling could be applied to NIH laboratories. "There is a possibility that the finding of the court in regard to DoD could be used as a precedent in challenging the NIH," says Robert McKinney, director of NIH's safety division. Although NIH has recently finished its draft environmental policy, it does not plan to automatically conduct environmental assessments for experiments with biological toxins. Spokeswoman Anne Thomas says the agency "is studying the case with some interest and discussing [possible responses] with the Department of Justice."

BL-4 materials include organisms not normally found in the United States with the potential of being highly infectious in humans; work on such organisms is done in sealed containers with gloves attached to the walls. BL-3 materials are generally organisms found in the country, such as tuberculosis, and can be safely contained within laboratory air hoods.

Christopher Anderson

Will less actually be more?

Bethesda, Maryland

HOPING to simplify the tangled system by which US universities account for their government-related research overhead expenses, officials are considering across-the-board limits that could return more than \$65 million per year to biomedical research.

At a meeting last week, an advisory panel to Bernadine Healy, director of the National Institutes of Health (NIH), proposed three different options for major reforms. All of the alternatives would cut out much of the accounting in favour of simple formulas or fixed limits, something that reflects the general opinion that the indirect cost overcharges are due more to the complexity of the system than to conspiracy. Notably, all the options (see box, this page) would reduce the amount that universities can charge as research overhead, on the argument that most universities would be willing to sacrifice a little reimbursement if they could eliminate the accounting morass.

Since the indirect cost scandal broke last spring, administration officials have been racing to put reforms in place, lest Congress do it for them. In May, the administration imposed a 26 per cent limit on the administrative portion of indirect costs as a stop-gap measure, promising to do more after further analysis. At the time, officials said this cap would save the government more than \$100 million a year (see *Nature* 350, 642; 25 April 1991). But the latest accounting, released last week, estimates that the saving will only be about \$55 million. NIH pays out about \$1,600 million each year for indirect costs.

Congress will be looking closely at the President's 1993 budget request for evidence of more substantial and lasting reforms. As that document is expected around the end of January, administration officials have just a month or so to reach consensus on new procedures. NIH are

just one of several agencies participating in the reform process, but because NIH have the largest slice of the research pie, Healy's recommendations are expected to carry some weight.

At the moment, Healy and many of her advisers appear to favour the option of a simple formula or a flat rate to determine costs. One of last week's proposals is to set either an allowable indirect cost rate based on the size, type and geographic location of a university, or else a flat rate of 48 per cent, across the board. At present, universities average a 51 per cent indirect cost rate, based on rates at each university that are arrived at by arcane calculations and long negotiations with government auditors. The advisory panel argued that eliminating that process should save much of the accounting overhead of research and would compensate for the cut in the overall rate, which would save \$65.6 million that could be used to fund 289 new research grants.

Whether this administration can be the one to finally fix the indirect cost system once and for all remains to be seen. University groups argue against wholesale change. "You really can't just scrap the whole thing and start fresh," says David Moore of the Association of American Medical Colleges. "You've got a huge research system out there that could collapse." Already many universities, squeezed by state budget cuts and tuition caps, are being forced to turn down research grants with low indirect cost limits (such as those from the Department of Agriculture, which are limited to 14 per cent) because they cannot afford to cover the overhead, he says.

But others argue for a shake-up all the same, despite the risk of teething pains. They point out that, unlike government contracts, where institutions do what the government wants in return for a fixed fee, research proposals come from univer-

NIH's options

- **Simplified accounting.** Universities that do less than \$3 million a year in government research are now allowed to use a short form to calculate their indirect cost rate, without elaborate accounting and negotiating. That cut-off could be increased to \$10 million, which would allow many more universities to go the 'EZ form' route. Potential savings: \$72 million.

- **Formulas or fixed rates.** Rather than negotiate rates on a case-by-case basis, universities could use a common formula based on their size, geographic location and public or private status. Or the government could do away with the whole process and go for a fixed 48 per cent rate, slightly under the current average. Potential savings: Between \$13 million and \$44 million depending on the formula used, or \$66 million for a fixed rate.

- **Caps.** Currently only the administrative portion of indirect costs is capped, at 26 per cent. That limit could be reduced to 20 per cent, library costs could be limited to two per cent, and student services charges could be eliminated. Potential savings: \$117 million. **C.A.**

sity scientists and benefit the institution as much as the government.

Indeed, the idea that universities deserve 'full-cost recovery' actually dates from wartime emergency rules, says John Holmfeld, a long-time congressional staff member who is now executive director of the Council of Scientific Society Presidents. Before the Second World War, universities were usually not given extra money to cover the incidental expenses of doing government-funded research. But during the war, the government asked universities to do special research in the national interest, and agreed to cover the extra costs. After the war, "it never stopped." **Christopher Anderson**

The case of the amazing shrinking scandal

Washington

REMEMBER the reports of tens and even hundreds of million dollars that US universities had improperly charged to research overhead? *Nature* reported those figures (350, 182; 21 March 1991 & 351, 173; 16 May 1991) and so did nearly everyone else. But last month government investigators released the results of a year-long audit and the problem appears to have been exaggerated — by a factor of ten or more.

The audit of 14 major US research universities found a total of between \$1.9 million and \$2.4 million in costs that universities had improperly charged

the government in the year examined. This is out of \$1,733 million in government research, or a bit more than 0.1 per cent. By accounting standards, that is not much. "Maybe it initially looked like a scandal," says Howard Gobstein of the Association of American Universities, "But we're finding out that the [accounting] system actually works pretty well."

So where did the big numbers come from? Part of the answer can be found in the complexity of indirect cost accounting. Just because a university may have improperly charged vast sums to government research accounts does not mean that it actually billed the government for

those amounts. Only a portion of the charges will fall under the category of billable research, and only a portion of research funding can be recovered as overhead. In last month's audit, investigators found about \$26 million of improper costs. But about half of that was voluntarily withdrawn, and most of the rest was not research-related. That left only \$3.2 million classified as research-related. Since each of the universities in the audit was given an allowable indirect cost rate of between 60 and 75 per cent, they actually overcharged the government for only about \$2 million total. **C.A.**

Anatomy of an accident

Tokyo

HUMAN error, defectively made technology, inadequate inspection procedures and lack of instructions for dealing with emergencies lay behind Japan's worst nuclear power accident, which occurred at the Mihama nuclear power plant last February, according to a final report on the accident released on Monday (24 November) by the Ministry of International Trade and Industry (MITI).

The Mihama accident, which marked the first time that an emergency cooling system of a Japanese reactor was triggered to prevent a meltdown of the reactor core, has shattered public confidence in nuclear power in Japan. And, by aggravating public resistance to nuclear power, the accident has helped to bring siting of new nuclear power plants to a halt at a time when the government is ambitiously trying to expand Japan's nuclear power base. The MITI report will do little to restore public confidence.

The accident occurred on 9 February when one of thousands of tubes in the primary cooling system of the No. 2 pressurized water reactor at Mihama burst open and more than 50 tons of radioactive primary cooling water flooded into the secondary cooling system. The emergency cooling system then automatically released a similar amount of water into the primary cooling system to prevent exposure and overheating of the reactor core.

A fundamental error in the manufacture of the cooling system more than 20 years ago and subsequent inadequacies in safety inspection procedures led to the burst in the primary cooling system, the MITI report concludes. The contractors

who assembled the boiler incorrectly installed anti-vibration bars that are supposed to suppress vibration of the thin U-shaped primary cooling tubes as high-temperature radioactive water rushes through them under pressure. In some parts of the cooling system at Mihama, the vibration bars were 40 to 50 cm shorter than required in the design, and, as a result, the cooling tubes in that area were left free to vibrate; one of them eventually sheared owing to high frequency fatigue at the point where the tube is held by a support.

The error in construction was not detected during inspections of the plant before commissioning nor in subsequent regular inspections of the plant over the past 20 years (nuclear power plants in Japan are shut down for several months every few years for safety checks). Damage to the tubes by excessive vibration was evident in eddy current tests of the tubes that were made during the regular inspections to detect holes in the tubes, but the evidence of fatigue was found only upon reexamination of the eddy current records after the accident. A MITI spokesman says the damage was not detected earlier because "no one ever imagined the anti-vibration bars were incorrectly installed" and so no one was specifically looking for fatigue near the tube supports.

A re-inspection of eddy current records from all 18 pressurized water reactors in Japan shortly after the Mihama accident revealed similar defects in the anti-vibration system of the nearby Takahama No. 2 reactor, and this reactor has also been shut down. Both the Mihama and Takahama reactors will be out of operation for several years while their steam generators are replaced at a cost of hundreds of millions of dollars.

But the problems at Mihama were not confined to the primary cooling tubes. After the leak occurred, the main steam isolation valve of the damaged steam generator failed to close properly in an automatic operation because over a period of time grease had leaked onto the sliding surface of the valve stem and turned into a "very sticky degenerated substance" that prevented full closure, the report says. Instead, the jammed valve was closed manually.

On Monday, MITI instructed electric utility companies to tighten up their regular inspections of the primary cooling systems of nuclear power plants, in particular the valves and anti-vibration bars. The ministry has also called on the companies to revise their operation manuals because the manuals at present "do not contain instructions for operating procedures in the event of a malfunction of the equipment".

David Swinbanks

New sheriff in town?

Washington

It may not be a fearsome investigative operation like that of Representative John Dingell (Democrat, Michigan), but the investigations and oversight subcommittee of the House of Representatives Science, Space and Technology (SS&T) committee is starting to show some of the same teeth. Taking a page out of Dingell's book, the subcommittee last week threatened a subpoena and a confrontational hearing if the Department of Energy (DOE) did not reveal its internal cost estimate for the Superconducting Super Collider (SSC).

On 20 November, the day before the subpoena was to be served, DOE backed down and agreed to let members of the subcommittee examine the documents. Although the issue of the SSC's cost — which has doubled since the project was first proposed — will be the subject of scrutiny, subcommittee staff say that they are not looking for a smoking gun. "We haven't had a tip that there's something nasty going on," says staff member Robert Roach. "The issue is the ability of Congress to exercise its oversight authority."

This is a notable shift for a committee that has traditionally been one of science's strongest supporters. Because the committee, which oversees the National Science Foundation and the National Aeronautics and Space Administration, controls no money, it has had little real clout in the past. But Dingell's investigations committee, which oversees the National Institutes of Health, also has no funding authority, but has managed to become a formidable force in shaping science policy, using nothing but the power of subpoenas and Dingell's intimidating cross-examinations. Although the SS&T investigations committee does not intend to take on scientific misconduct soon, it has shown that it is willing to investigate many of the research projects it once held sacred.

In the case of the SSC, the investigations committee has been requesting internal memoranda between SSC programme director Joseph Cipriano and DOE secretary James Watkins giving periodic updates on the project. Although DOE provided the subcommittee with some 40 boxes of other SSC documents, the department first refused to acknowledge that the Cipriano documents had ever existed, then said that they might have been destroyed. After the subcommittee pressed the issue, DOE admitted the documents did exist, but said it would not reveal them, citing as a precedent President Richard Nixon's refusal to turn over documents relating to the Watergate scandal. Now the DOE has given in on that point too, and the subcommittee intends to investigate the possibility that DOE has systematically underestimated the cost of the SSC to avoid congressional opposition. Christopher Anderson

ENVIRONMENTAL POLICY

GATT attack

Washington

Two US congressmen introduced legislation last week to prevent the Administration from weakening US environmental laws to protect international trade. The bill, by Henry Waxman (Democrat, California) and majority leader Richard Gephardt (Democrat, Missouri), warns the Administration that Congress will not stand for the 'harmonizing' of US environmental rules simply to conform with international standards, a threat that first arose in August, when negotiators for General Agreement on Tariffs and Trade (GATT) talks decreed that countries could not allow their internal environmental policies to get in the way of international trade. At the time, GATT ruled that the United States had violated the trade agreement when it banned Mexican tuna imports on the grounds that Mexican fishing fleets kill too many dolphins.

C.A.

Translations and turkey

New York City

TWENTY-THREE editors of Soviet physics journals assembled in New York last week under the auspices of the American Institute of Physics (AIP) to meet their US counterparts and discuss the translation of Russian-language journals into English and vice versa. At the end of the meeting, over a traditional US Thanksgiving Day meal of turkey, dressing and cranberry sauce, the participants agreed that the contacts they made would serve them well in bringing the two communities closer together, and, in particular, that the US scientists were given valuable insight into the problems now facing Soviet science.

The AIP has contracts to translate 20 of the best Soviet physics journals into English, including *Doklady* and the *Journal of Experimental and Theoretical Physics*. Recently, the institute has begun to publish some of its translations simultaneously with the release of the Russian-language journals by obtaining the contents of the journals before they are published. Indeed, Valerii Ozhogin, editor of *Superconductivity*, said he received the English translation of the most recent issue of his journal before he got the Russian version. This has given him ammunition, he said, to prod the publisher of the journal to speed up its publication.

Although the US representatives reported some complaints and concerns about the quality of the translations, the Soviets for the most part seemed satisfied with that part of the arrangement. They did not, however, seem ready to make any guarantee that the AIP would remain the exclusive translator of the physics journals. That translation is a lucrative business, and some of the Soviet editors seemed to sense that they might get a better deal by opening up the bidding.

The most sensitive issue was whether the Soviet journals would remain Russian-language publications or would change to English, as have the top scientific journals in continental Europe and Japan. The consensus among the Soviets seemed to be that with successful simultaneous publication of English- and Russian-language versions of the journals, there is little reason to switch to English, and that there are good reasons not to.

Many Russian scientists do not read or write English well, Ozhogin said, and they would be seriously handicapped if the best Soviet papers were published in English. Few English-language journals are translated into Russian editions, however, so Soviet scientists must find some way to get around the language barrier. Apparently the scientists whose specialties make

it essential for them to read English-language journals do learn to read English, but the majority of Soviet scientists are still uncomfortable with the language.

The Soviets acknowledged that more and more of their best researchers are getting jobs in the West, particularly with the mounting economic problems in the Soviet Union. This, combined with the general opening up of the country has led to more Soviet scientists publishing their papers in English, and this puts pressure on the Russian-language publications. *Superconductivity* accepts manuscripts in English and translates them for publication, for instance, but even so there are worries that English-language publications will skim off the cream of Soviet science, leaving the Russian-language publications with the lesser papers.

One fact came out at the meeting that illustrated how difficult it is for Soviet scientists to move around and communicate with their peers. Most of the Soviet editors were meeting each other for the first time — they had to travel to the United States to meet colleagues who live in different cities in the Soviet Union.

Another discovery was that the Soviets are just as confused about what to call their country as the Americans. One US speaker stopped in mid-sentence to ask, "Is it the Soviet Union or Russia?" The response from one visiting scientist: "We don't know either."

Robert Pool

ELECTROMAGNETIC FIELDS

Little accord on priorities

Washington

THE US government is apparently becoming more serious about investigating the claims that low-frequency electromagnetic fields (EMFs) may affect human health. Last week, officials from the Department of Energy (DOE) assembled EMF experts from universities, electric utilities, and state and federal agencies to begin developing a coordinated federal programme to study possible connections between EMF exposure and increased risks of cancer and other health problems.

In the 1992 appropriations bill that set the DOE budget, Congress stated that it supports "continued research on the potential health effects of electromagnetic fields" and asked "that the Department of Energy be the lead agency for such research and that research conducted by the government be coordinated through the Department." Congress did not, however, increase its funding for EMF research in fiscal year 1992 — the DOE's portion, for instance, stayed approximately constant at \$5 million.

The energy department moved quickly to begin coordinating EMF studies, which are supported not only by DOE but also by a variety of federal and state agencies and

utilities. In the past, what little coordination that existed among these programmes has been informal and sporadic.

The organizers of the DOE workshop had hoped to set some priorities on EMF research, but with the wide range of researchers represented, every approach to studying the problem had its defenders. Everyone agreed, for instance, that it is essential to develop some idea of what 'dose' means when speaking of exposure to EMFs — most research indicates that it is not enough to measure simply the magnitude and duration of the applied field, but factors such as frequency, orientation, peak field and pulsing may also be important. However, some participants argued from this for more laboratory work, while others urged better epidemiological studies. The upshot was that no clear priorities were evident at the end of the meeting that were not already generally accepted at the beginning.

Nonetheless, DOE intends to keep to an ambitious schedule for establishing its unified EMF strategy. It plans to develop the strategy over the next month, hold public meetings to discuss it in January and February, and begin to implement the plan in the spring.

Robert Pool

UNDERWATER EXPLORATION

Jason lost at sea

JASON Jr., the robotic underwater explorer that first explored the *Titanic*, was lost last week when the barge carrying it to the Galapagos islands foundered and sank in 9,000 feet of water off Ecuador. No lives were lost, but the submarine and several containers of video and electronic equipment are considered unsalvageable. Most of the \$20 million loss was insured, says Woods Hole Oceanographic Institution spokeswoman Shelly Lauzon.

Early this week, Jason officials said they were "optimistic" that the project — which is to provide a glimpse of a live scientific expedition to 500,000 students in North America — will be able to begin on 2 December as planned. Woods Hole and the Massachusetts-based Jason Foundation for Education have located three available submersibles and replacements for much of the electronic equipment. They hope to airlift the supplies to the Galapagos by mid-week. The replacement submersibles, known as mini-rovers, are designed for commercial uses in much shallower water than Jason Jr., which can descend to depths of several thousand feet. But because the project is only designed to explore the islands at depths of a few hundred feet, team members hope to be able to stick to their original plans. C.A.

Crisis at London Zoo

SIR — London Zoo is in crisis. It has operated at a loss for a decade. The zoo's executives, the council of the Zoological Society of London (ZSL) who put them in place, and the society's fellows can all agree on that. Closure has been threatened, and used to menace the public (with some success) and the government (latterly unsuccessfully) into giving money to "Save Our Zoo". At the society's recent annual general meeting and in two subsequent meetings arranged for management to air its plans, many fellows have been confirmed in their belief that the crisis is more deeply rooted than in finances alone. There have been numerous failures to manage well and to negotiate for funds appropriately, but in a sense these are symptomatic, not causative. Rather, the zoo has drifted for years — and has now drifted near to a decision to close — without either defining for itself a role which is self-evidently valuable, or having any guiding philosophy. It is failing precisely because it is seen to have neither. We fear that if the zoo closes, the Institute of Zoology, and perhaps the society, will follow shortly afterwards.

We think structural change in management is needed. The present executive has hawked grand plans for theme park

developments since 1988. A low-cost attempt to adopt straightforward conservation ideals such as those of the World Conservation Union (IUCN), backed by serious education and research, does not seem to have occurred to them. We contend that this worthwhile purpose is likely to ensure the future of London Zoo through visitors and sponsors just as it currently secures more enlightened zoos worldwide. Some seven directors run the business, yet there is no full-time Director of London Zoo to embrace this kind of thinking.

With the aim of reorganizing management to the ultimate benefit of the zoo, we and a hundred or so other fellows are "requisitioning" a special general meeting of the society to debate resolutions to this end. (The terminology, and our very restricted opportunities for action, derive from our early nineteenth century charter.) Quaintly, but frustratingly, no list of members' addresses is allowed to us. In consequence we shall be grateful if ZSL fellows and others wanting more information would contact us at the address below.

CHRISTINE HAWKEY, RUTH SOBER, SIMON
BEARDER, STEPHEN COBB, COLIN TUDGE,
GRAHAM MITCHELL

Ad Hoc Group of ZSL Fellows,
13 St Giles, Oxford OX1 3JS, UK

Continental drift

SIR — The hypothesis of continental drift and its subsequent generalization to plate tectonics has revolutionized geological thinking. Although Wegener championed the cause of continental drift, he was not the first to suggest that the continents moved apart. Hallam¹ points out that we must distinguish between the case where only similarities between continental coastlines are noted and the case where lateral motion is also suggested. Antonio Snider-Pellegrini (referenced by Hallam) did suggest continental break-up and drifting in 1858 in *La Création et ses mystères dévoilés* but I recently discovered a publication that predates this work by 20 years.

Thomas Dick² of Broughty Ferry, Scotland, in a book entitled *Celestial Scenery; or The Wonders of the Planetary System Displayed; Illustrating the Perfections of the Deity and a Plurality of Worlds* and, according to Houzeau originally published in 1838³, wrote:

There is a striking correspondence between two sides of the two continents to which we have adverted, the prominent parts of the one corresponding to the indentings of the other. If we look at a terrestrial globe or map of the world, we shall perceive that the projection of the

eastern coast of Africa nearly corresponds with the opening between North and South America, opposite to the Gulf of Mexico; that the projection in South America, about Cape St. Roque and St. Salvador, nearly corresponds with the opening in the Gulf of Guinea; so that, if we could conceive the two continents being brought into contact, the openings to which I have referred would be nearly filled up, so as to form one compact continent A consideration of these circumstances renders it not altogether improbable that these continents were originally conjoined, and that at some former physical revolution or catastrophe, they may have been rent asunder by some tremendous power, when the waters of the ocean rushed in between them, and left them separated as we now behold them.

Unless an earlier publication by someone else is found, Dick deserves credit for being the first specifically to mention the possibility of continents' moving apart.

ALAN GOODACRE

Geological Survey of Canada,
1 Observatory Crescent,
Ottawa, Ontario,
Canada K1A 0Y3

1. Hallam, A. *A Revolution in the Earth Sciences*, 127 (Clarendon, Oxford, 1973).
2. Dick, T. *Celestial Scenery*, 181 (reprinted by W. P. Fetridge and Co., New York).
3. Houzeau, J. C. & Lancaster, A. *General Bibliography of Astronomy to the Year 1880*, Vol. II, 2225 (reprinted by Holland, London, 1964).

Worrying trends

SIR — It would be difficult to fault much of Evelyn B. Sherr's defence of prophets of catastrophic outcomes of present trends of human development (*Nature* 352, 752; 1991). I wonder, however, if a continuation of Sherr's own arguments might not give a more optimistic picture.

I can readily accept that no plausible extrapolation of present trends would give to the 1988 world population of 5,000 million, still less the projected AD 2100 population of 15,000 million, the \$11,275 gross national product (GNP) per head of the more favoured developed nations. If, however, such an outcome is beyond the resources of the planet, it will not happen, so we need not get too worried about it. Nor is it at all obvious that those nations that do not acquire their electric toothbrushes and bottled water will be missing all that much.

Moreover, as far as the developed industrial countries are concerned, there is no evidence to suppose that their continued development will be restricted by the resources of the planet in the foreseeable future. Far from running out of soil, they are forced to take land out of production to avoid unmanageable food surpluses. At some time, they will exhaust fossil fuel supplies, but when they are faced with an unavoidable choice they will resort to nuclear energy to whatever extent is required to stop the lights going out. Nor does population growth appear much of a threat in more developed countries.

One is then left with Sherr's fear about the consequences of unchecked population growth in less developed countries. Here Sherr would not have an easy task to show that human communities have historically tended to grow beyond the resources required to sustain them. Moreover, if they did, there would be little point in the developed countries providing resources for them, since further population growth would presumably exhaust these as well.

Perhaps, if left to themselves, the pre-industrial countries will solve their own problems. The industrial countries should provide neither weapons to facilitate their genocidal civil wars nor the promise of long-term aid that would lead them to suppose they have no need to worry about their own population pressures.

If this appears uncharitable, it is surely less so than a policy for developed countries that deliberately sustains population growth in developing countries to a point where the problem is beyond all remedy.

ALFRED ROUT

73 Cambridge Road,
Middlesbrough,
Cleveland TS5 5NL, UK

AIDS, monkeys and malaria

Charles Gilks

Could primate retroviruses have been passed on to man or other monkeys as a result of experiments with primate malarias? An answer to this question could explain the origin of the AIDS epidemic.

SEVERAL theories have been proposed to explain the origins of the AIDS epidemic^{1,2}. Most have been speculative rather than testable and several have caused offence³, particularly those which refer to sexual practices with monkey blood. The important question of the origin of human immunodeficiency virus (HIV) awaits a satisfactory answer.

Three separate outbreaks of primate lentiviruses have been recognized: HIV-1 in Europe, North America and central Africa; HIV-2 in West Africa; and simian immunodeficiency virus (SIV_{MAC}) in North American primate colonies. Both HIV-1 and -2 may have been endemic in parts of Africa for centuries and become epidemic only with the recent profound social and economic changes in these areas. Alternatively, both viruses may have recently entered the human population, in which case the likely origin of HIV-1 is the closely related SIV of chimpanzees, SIV_{CPZ} (ref. 4), and that of HIV-2 is the retrovirus of sooty mangabeys, SIV_{SMG} (ref. 5). There is little doubt that SIV_{MAC} was originally SIV_{SMG} or a close relative^{5,6}. The mechanisms of cross-species transfer proposed^{1,5} do not offer a plausible unifying hypothesis to explain the three outbreaks.

Direct inoculation of fresh blood is the most efficient way to transmit the AIDS virus. No one has suggested any circumstances under which fresh monkey blood could have been injected into humans in a systematic fashion, and this has not been considered as a mechanism. But the malaria literature describes many instances in which humans were injected with primate blood containing viable malaria parasites^{7,8}. Most interest centred on plasmodia of Asian primates, none of which is thought to harbour SIV-like retroviruses naturally⁹. But humans have been directly inoculated with chimpanzee or mangabey blood, and macaques with mangabey blood.

In 1922, Blacklock and Adler injected themselves with fresh chimpanzee blood infected with *P. reichenowi*¹⁰ to see if humans were susceptible to the parasite which might therefore be zoonotic *P. falciparum*. In 1939, Rodhain documented a further 12 attempts to transfer *P. reichenowi* to man¹¹. Five chimpanzees were used and the volumes of blood transferred ranged from 2.5 to 15 ml. Most of the chimpanzees would have

come from the Belgian Congo, now Zaire. The chimpanzee also has a quarantain parasite, *P. rodhaini*, now considered synonymous with *P. malariae*. In various experiments, 6 people were given chimpanzee blood and a further 27 patients with neurosyphilis were infected in sequences of up to 14 human-to-human passages^{11,12}.

P. schwetzi is the tertian parasite of the chimpanzee, and has close affinities with both *P. vivax* and *P. ovale* of humans. Rodhain describes ten humans challenged with chimpanzee blood infected with *P. schwetzi*^{11,13}. In 1954–55, a further four adults were challenged with fresh, parasitized chimpanzee blood. Several of these people were then used as the source of blood given to a further six adults¹³. *P. schwetzi* has also been given to volunteer prisoners in the United States, by mosquito passage and direct human-to-human passage¹⁴.

Fewer experiments have been done with *P. gonderi*, the tertian parasite of sooty and agile mangabeys (*Cercopithecus atys* and *C. galeritus*) and the drill (*Mandrillus leucophaeus*). Coatney mentions two attempts to transmit this to man by direct inoculation of mangabey blood¹⁵. Rodhain and van den Berg infected a series of primates with *P. gonderi* originally obtained from an agile mangabey. Two humans were given blood from one macaque with a patent infection and a further neurosyphilitic patient was given parasitized blood from another infected macaque¹⁶. More recently, *P. gonderi* has been transmitted by blood passage to 17 rhesus macaques (*Macaca mulatta*) and to an unspecified number of *M. radiata*⁸.

Thus at least 34 people have received parenteral injections of fresh blood taken from 17 chimpanzees. A further 33 received blood from people given the primary chimpanzee blood injections. Far fewer, perhaps two, are described as being given direct inoculations of mangabey blood. In addition, three people have received blood from macaques infected with mangabey malaria parasites passaged via a baboon. All these recipients of primate blood must be considered at risk of developing retroviral infections if the hosts were infected with SIV_{CPZ} or SIV_{SMG}, respectively.

It is possible, then, to propose a unifying theory to describe how primate retroviruses may have crossed species boundaries into man or other monkey

species as an unforeseen hazard of primate malaria experiments. The data are by no means conclusive. There are only a few documented instances of humans having been given mangabey blood. Most of the experiments were in Europe, whereas HIV-2 probably started in West Africa. Some experiments may not have been published, particularly if they yielded negative results. The literature review may be incomplete.

But my theory is testable. It should be possible to determine more accurately the exact background to the experiments in the United States in which macaques may have been given mangabey blood, and to establish what happened to these monkeys after the project was over. The records of human volunteer experiments, particularly those involving prisoners in the United States⁸, can be checked to see if any further, unpublished studies took place. Material from several of the original experiments could still exist and could be tested for the presence of retroviruses. Serum samples from some of the patients or from the original primates could have been stored, and Giemsa-stained slides could be available in slide archives. Some of the primate malarias may be cryopreserved and could even contain recoverable retroviruses. It is to be hoped that those with access to this material will test my hypothesis so that the debate about the origin of AIDS can become more scientific. □

Charles Gilks is at the Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK and is a visiting scientist at the Kenya Medical Research Institute, PO Box 43640, Nairobi, Kenya.

1. McLure, M. *New Scientist*. 54–57, 30 June 1990.
2. Karpas, A. *Nature* **348**, 578 (1990).
3. Owusu, S. K. *Nature* **350**, 184 (1991).
4. Peeters, M. et al. *AIDS* **3**, 625–630 (1989).
5. Hirsch, V. M. et al. *Nature* **359**, 389–391 (1989).
6. Desrosiers, R. C. et al. *AIDS Res. Hum. Retrovir.* **5**, 465–473 (1989).
7. Garnham, P. C. C. *Malaria Parasites and Other Haemosporidia* (Blackwell, Oxford, 1966).
8. Coatney, G. R. et al. *The Primate Malarias* (US Department of Health, Education & Welfare, 1971).
9. Schneider, J. & Hunsmann, G. *AIDS* **2**, 1–9 (1988).
10. Blacklock, B. & Adler, S. *Ann. Trop. med. Parasit.* **16**, 99–106 (1922).
11. Rodhain, J. *Ann. Soc. Belge Trop. Med.* **19**, 563 (1939).
12. Rodhain, J. & Dellaert, R. *Ann. Soc. Belge Trop. Med.* **23**, 19–46 (1943).
13. Rodhain, J. & Dellaert, R. *Ann. Soc. Belge Trop. Med.* **35**, 757–775 (1955).
14. Contacos, P. et al. *Am. J. Trop. med. Hyg.* **19**, 190–195 (1970).
15. Coatney, G. R. *Am. J. Trop. med. Hyg.* **17**, 145 (1968).
16. Rodhain, J. & van den Berg, L. *Ann. Soc. Belge Trop. Med.* **16**, 521–531 (1936).

Where has fundamental physics gone?

Can an excess of success in solving problems have induced physicists to overlook the serious intellectual difficulties that surround them?

In Aristotle's time, nearly 3,000 years ago, there was still everything in science to play for. To have the wit to ask what sustains the motion of a projectile, say a thrown stone, was a triumph in itself. Aristotle's explanation that a projectile will create a void into which air will rush, impelling the projectile on its course, has the merit of seeming consistency, but no other. Yet, by drawing attention to conceptual issues, it did much to keep succeeding millennia on their toes.

So, for that matter, did Einstein, with $E = mc^2$. Since then, of course, there have been general relativity, quantum mechanics, nuclear structure, the recognition of Hubble motion in an expanding Universe, quantum electrodynamics more generally, and even ambitions that all four forces of nature (if that is how many there are) will be unified. So why has physics as practised (or, at least, as published) boiled down to mere problem-solving?

Aficionados will deny the slur, but they will be wrong. Those who even occasionally look at this space will have been given the impression that there is no better guide to what is interesting and important in physics than *Physical Review Letters*, the weekly journal of the American Physical Society. The impression is correct. But the grand themes of Aristotle and Einstein seem to have given way to declarations such as "arrays of coupled limit-cycle oscillators are of fundamental importance in physics, biology and engineering" and "The existence of isospectral γ sequences in a number of neighbouring superdeformed nuclei poses an exciting challenge to nuclear theorists". Each statement is the first sentence of an article in *Physical Review Letters* for 11 November, chosen at random. Where are the grand themes of the old days? And why have they disappeared? To be fair, nobody in his or her right mind would expect that every paper in a journal running to more than 140 pages a week would raise a fundamentally novel question. But should not some of them do that?

Pedestrianism, sadly, seems to have taken over. So much can be gathered from the way in which people sidle up to the question of whether there can be a correspondence between the stable states (eigenstates) of quantum systems and the states in which the corresponding classical systems are prone to what is called 'deterministic chaos'.

Evidently it would be a question of

Aristotelian grandeur to ask whether the correspondence is invariable and in which directions the words 'necessary' and 'sufficient' apply. But the practitioners seem more eager to solve problems, sometimes reporting how much CPU time they have used in their numerical solutions. We (the readers) are still awaiting news of whether classical chaos means an eigenstate, or an eigenstate means classical chaos, necessarily or sufficiently as the case may be.

The disappointments seem widely spread. In the previous issue of the same journal, more than 200 people describe an experiment at Fermilab that failed to detect the decay of charged heavy vector bosons (elaborations of the intermediate vector bosons $W^\pm$ and Z^0) into leptons, leading them to conclude that such particles will not be found with any accelerator now being built. And a small army of Japanese showed that the decay of the strange K^0 meson into positive and negative pions is so much more common than their decay into muons and electrons that these processes can be neglected as literally insignificant. Nobody will be surprised; it would have been different if the experiments had turned out differently.

Understandably, in the circumstances, there seems a great deal of fine tuning under way. In an earlier issue, there is a scholarly but inconclusive discussion of whether there are neutrinos with a mass of 17 keV and some experiments that show that highly excited caesium and rubidium atoms (in Rydberg states) behave as if they were chaotic systems.

Little things irritate, and are bound to do so, when there is so little grand to say. But it is more than a passing inconvenience that readers should be told as often as they are that a crucial reference has, as yet, no citation, but only the phrase "to be published". It is especially mean that this should apply to pieces of algebra.

Inevitably, what catches the eye (and sticks in the mind) during times of ennui such as this is experimental neatness. Luckily (to judge from *Physical Review Letters*) a lot of that persists.

Take, for example, the case of the group at the University of Colorado at Boulder that has devised a novel way of storing particles (such as atoms) that have a magnetic moment. Using the absorption lines of, say, caesium, atoms are first 'cooled' by means of six off-time laser beams (one from each face of a surrounding cube) in what is now a relatively familiar trap of

'optical molasses'. When the optical trap is filled, the atoms are then propelled vertically upwards, through exactly 14 centimetres, into what is nothing but an axially symmetric bottle oscillating between axial and radial confinement so that the atoms are further 'cooled', or robbed of kinetic energy. The trap is an obvious means of dealing with spin-polarized atoms. The group promises next to measure something.

There are also several neat ideas put into circulation. What happens, for example, to an atom adsorbed on a crystal-line surface of identical composition? STM measurements have now made it clear that even apparently perfectly flat surfaces are marked by ledges, just one atomic diameter high, where single layers of atoms come to an edge. An atom adsorbed below a ledge will, by diffusion, eventually unite with it, but one adsorbed above a ledge is much less likely to tumble over, and so will be a roughening influence.

The same issue of the same journal has an account of a humble study of the aggregation of polar molecules on a water surface. Circular patches become hexagonal when the aggregation rate is right. The explanation has to do with the lesser surface free energy of a shape whose curvature is concentrated at the corners. Is that why snowflakes are hexagonal?

Interesting though it may be, all this is the small change of physics, hardly worthy of the attention of an Aristotle or of an Einstein. Physics is going through a dull patch, when even predicted attainments, such as the discovery of the top quark, will merely bring relief (if they are made) and a clamour for bigger accelerators (if they are not).

Although it would be irresponsible to suggest that physics should manufacture ideological controversy to keep itself in the public eye, the more serious danger is that its *élan* may be destroyed by the complacency that problem-solving brings. And that is an odd conclusion from too much reading of an excellent journal.

John Maddox

Correction

LAST week's News & Views leading article (*Nature* 354, 183; 1991) attributed to Joseph Goldstein the astounding allegation that half of all cytoplasmic proteins may be prenylated: the figure he actually named was 0.5 per cent. It should also have referred to lipid prenyl tails, not lipid tails. Our apologies to Dr Goldstein and to our readers.

Getting under the skin

Michael W. Klymkowsky

STRIKING news for students of human disease and intermediate filaments comes in a series of five reports¹⁻⁵ from three different groups. Not only do they identify the underlying cause of epidermolysis bullosa simplex (EBS), one of several inherited disorders characterized by blistering of the skin, but they also point to just what keratin, the type of intermediate filament found in the epidermis, actually does.

Intermediate filaments are cytoplasmic skeletal proteins that are found in many different types of cell. The keratins are the most diverse members of the family, numbering about 30 different proteins, and there are two principal forms of them: the β , which form β -pleated sheets, and are found in silk and the claws and beaks of birds, and the more widespread α , characteristic of skin, hair and nails. The α -keratins are subdivided into two groups, type I and type II (see ref. 6 for review). Neither can polymerize into a filament on its own, but any type I and type II keratin together can do so.

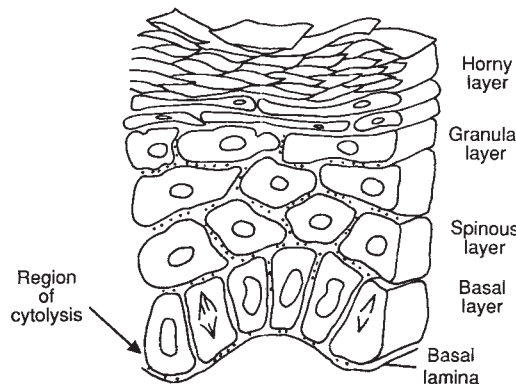
The function of intermediate filaments is thought to be basically structural, but when their organization is disrupted in cultured cells, there is little if any noticeable effect on cellular behaviour⁶. So researchers have had to look to more complex systems.

Three of the papers discussed here^{1,3,5} are from a group led by Elaine Fuchs, who set out to exploit the discovery⁷ that truncated forms of the human K14 keratin-type protein act as dominant antimorphic⁸ or negative-effect⁹ mutations, disrupting endogenous keratin-filament organization. The K14 protein is the main type I keratin expressed in the basal cell layer of the mammalian epidermis. By using six kilobase pairs of upstream K14 sequence, mutant genes were specifically expressed in the basal cell layer of transgenic mice. Many of the mice exhibited profound defects in the epidermis, ranging from its complete absence to severe blistering. These symptoms, together with ultrastructural features of the epidermal cells, were quite like those found in people with EBS¹. But it remained unclear whether the epidermal phenotype was due to the disruption of keratin filaments or to the formation of large, mutation-induced keratin aggregates.

To resolve this question, transgenic mice that expressed one of a series of keratin mutants were constructed³; the severity of the epidermal phenotype in mice correlated with the severity of the mutation's effects on keratin-filament

organization in cultured keratinocytes. A less severe mutation did not produce massive keratin aggregates yet still produced an EBS-like phenotype, suggesting that the disruption of keratin-filament organization was responsible for the phenotype.

Studies of EBS patients provided the final link between keratin mutations and the disease^{2,4,5}. Ito *et al.*² reported the absence of keratin filament bundles in



Structure of the mammalian epidermis. The mutations in EBS patients occur in the genes for keratins K5 and K14, which are expressed mainly in the basal layer. It is there that cytolysis mainly occurs.

the skin basal cells of one patient, while Bonifas *et al.*⁴ looked at both the pedigree and (using restriction fragment length polymorphism (RFLP) analysis) the genotype of two families with a history of EBS. They found RFLPs that cosegregated with the disease syndrome and were closely linked to the type II basal layer keratin K5 gene in one family and to the K14 gene in the other. Moreover, the K14-linked RFLP appeared in the same generation as the onset of the disease, further implying that here was a case of cause and effect. Sequence analysis of the K14 gene revealed a single base-pair change in which the leucine at position 384 was replaced with a proline. Finally, in the study of Coulombe *et al.*⁵, point mutations were identified in the K14 genes of two unrelated Dowling-Meara patients (Dowling-Meara is a subtype of EBS). Both involve the highly conserved arginine at position 125; in one the arginine is replaced by a cysteine, in the other by a histidine. The authors then re-created these mutations, introduced them into cultured epidermal keratinocytes and observed that they induced the disruption of keratin filaments.

In EBS, cytolysis leads to the separation of the suprabasal layer of the epidermis from the dermis, and to blistering. How could the disruption of keratin filaments result in this effect?

The mutations identified in the EBS patients occur in keratins (K5 and K14) that are expressed primarily in the basal cell layer. In the suprabasal layers, the synthesis of these keratins is low and other keratins are expressed in their place. Mutations in either K5 or K14 would therefore be expected to affect primarily the keratin filaments of basal cells and to a lesser extent those of the overlying spinous cell layer. The structure of the epidermis shows why cytolysis occurs primarily in the basal layer (see figure). At the basal surface the cells are anchored to the underlying connective tissue, which provides structural support. At their apical surface, basal cells are anchored to one another and to the cells of the spinous layer through desmosomes, the strong intercellular junctions that bind cells together. Because desmosomes are not affected by the disruption of keratin filaments⁶, much of the mechanical integrity of this region should be maintained. Nuclei, with their intermediate-filament-like nuclear lamina, could also provide some mechanical strength. This leaves a weak link, presumably normally strengthened by keratin filaments, between the apical and basal surfaces of the basal cell where cytolysis begins^{2,3}.

Further study of intermediate filaments is likely to rely heavily on the use of antimorphic mutations. But a word of caution is in order. Trevor¹⁰ constructed an embryonal carcinoma cell line that expressed a mutant keratin that disrupts keratin filament organization; her surprising observation was that these cells are unable to form normal epithelia when induced to differentiate into embryoid bodies. In contrast, previous studies had indicated that keratin filaments are not required for either forming or maintaining an epithelium⁶. Clarification comes from the work of Baribault and Oshima¹¹, who constructed an embryonic stem cell line homozygous for a disruption of the keratin 8/Endo A gene. This is the only type II keratin expressed in these cells, and keratin

1. Vassar, R., Coulombe, P. A., Degenstein, L., Albers, K. & Fuchs, E. *Cell* **64**, 365-380 (1991).
2. Ito, M. *et al.* *Arch. Dermatol.* **127**, 367-372 (1991).
3. Coulombe, P. A. *et al.* *Cell* **66**, 1301-1311 (1991).
4. Bonifas, J. M., Rothman, A. L. & Epstein, E. H., Jr *Science* **254**, 1202-1205 (1991).
5. Coulombe, P. A., Hutton, E. M., Vassar, R. & Fuchs, E. *J. Cell Biol.* (in the press).
6. Klymkowsky, M. W., Bachant, J. B. & Domingo, A. *Cell Motil. Cytoskel.* **14**, 309-331 (1989).
7. Albers, K. & Fuchs, E. *J. Cell Biol.* **105**, 367-372 (1991).
8. Muller, H. J. *Proc. 6th Int. Congr. Genet.* 213-255 (1932).
9. Herskowitz, I. *Nature* **329**, 219-222 (1987).
10. Trevor, K. T. *New Biologist* **2**, 700-711 (1990).
11. Baribault, H. & Oshima, R. G. *J. Cell Biol.* (in the press).

filaments cannot form in its absence. When induced to differentiate, these keratin 8/Endo A-minus cells normally form epithelia, meaning that mutations in keratin genes can, at least in some circumstances, do more than disrupt keratin filaments. Because most mutants are characterized in relatively simple cultured cell systems and then used in more complex differentiating or embryonic systems, the possibility of such unexpected effects must be addressed. In this sense, the phenotype of intermediate-filament gene disruptions should provide a framework within

which to judge studies of the mutant protein products.

Nevertheless, the new work on EBS provides the first real evidence for the mechanical functions of intermediate filaments. The real question now is whether they are strictly mechanical elements of cells and tissues, or whether they have some other, as yet undefined, purpose. □

Michael W. Klymkowsky is in the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA.

SPACEWATCH

Our asteroid-pelted planet

Duncan Steel

In January, the news media were full of stories concerning a record near-miss of the Earth by an asteroid. On page 287 of this issue, Scotti *et al.*¹ give the definitive view of their discovery of that asteroid (1991 BA), which sparked interest — and some alarm — around the world.

In September last year, after several years of development, the Spacewatch telescope at Kitt Peak in Arizona began to be routinely operated in a search for Earth-approaching asteroids. With the observing duties being shouldered by the principal investigator (Tom Gehrels) and the first two authors (Jim Scotti and David Rabinowitz) of the new paper, 18 nights of data collection per lunar month result in the discovery of typically one or two Earth-crossing asteroids, along with a host of other interesting objects, such as the periodic comet Spacewatch (1991x). This programme is the first to implement a search for such fast-moving objects using CCD (semiconductor charge-coupled device) detectors; the other three significant programmes (two run by Gene and Carolyn Shoemaker and by Eleanor Helin from Mount Palomar in California, and our own at the Anglo-Australian Observatory) use photographic techniques. The third author of the new paper, Brian Marsden, is responsible for the assimilation and sorting of data from all such searches, with near-Earth objects (NEOs) being found occasionally in other projects run from Japan, Chile, Europe and elsewhere.

The discovery of 1991 BA was remarkable in that it represents the closest observed approach, just 170,000 km, to the Earth by an asteroid (the Moon is

400,000 km from us). 1991 BA is also the smallest (about 5–10 metres across) and intrinsically faintest astronomical object ever observed telescopically. On 7 October, another asteroid (1991 TU), which is similarly faint (and perhaps, it turns out, the same size), was observed with the Spacewatch telescope over the course of two hours, passing about 750,000 km from the Earth². Although



Devastation caused at Tunguska, Siberia, on 30 June 1908 by an asteroid or 50-metre cometary nucleus that exploded in the atmosphere. Over 1,000 km² of forest was flattened and over a third of atmospheric ozone destroyed by the explosion. (Photo from Novosti.)

photographic searches still result in a large fraction of the discoveries of NEOs, CCD-based systems are clearly the preferred technique for the future.

Activity in this area seems certain to escalate. Under direction from the US Congress, NASA is currently studying ways of, first, detecting all asteroids and comets that threaten the Earth and civilization; and, second, diverting, destroying or otherwise ameliorating the effects of predicted impactors. The first task has been tackled by a committee (including myself) chaired by David Morrison, which reports in early January. The second is the concern of a team

being assembled by John Rather and due to meet in Washington in mid-January. At its recent general assembly in Buenos Aires, the International Astronomical Union (IAU) also set up a working group to consider the threat from such impacts.

In addition, the Soviet Union is attempting to stimulate international action on this front. On 10–11 October, the Institute for Theoretical Astronomy in St Petersburg hosted a conference, *The Asteroid Hazard*, attended by several Western astronomers and others in addition to a large home contingent. It was apparent that the Soviet programme is well advanced. Participants agreed that the available evidence indicates the frequency of impacts by 50-metre objects to be close to one a century, and that these may have catastrophic effects beyond those of the Tunguska event over Siberia in 1908 (see illustration).

In this regard it is interesting to hark back to the case of 1991 BA: if it had struck the Earth it would have been at a velocity of 21.2 km s⁻¹. With a size of about 8 metres and a typical meteoritical density, the explosive power would have amounted to around 40 kilotons of TNT, about three times the energy of the Hiroshima bomb. The 'annual event' for the Earth is apparently a 20-kiloton airburst³. Although the chance of an individual object like 1991 BA hitting the Earth is only 10⁻¹⁰–10⁻⁷ a year (depending upon its orbital parameters), there are believed to be 10⁹ such NEOs larger than 10 metres across. This implies that, each day, at least one passes by the Earth closer than did 1991 BA, but escapes detection. Asteroid 1991 BA was discovered only because the sole instrument capable of detecting and tracking it was pointed in the right direction at the right time, and because the operators were skilled and experienced enough to recognize and follow it.

To determine the orbit of a NEO accurately and to calculate whether it may strike the Earth in the foreseeable future, radar observations make a useful adjunct to optical tracking; in fact radar tracking for a small NEO on collision course with us could allow the point of impact to be determined, in principle, to within 50 km, and would thus allow evacuation. The guru of such radar work is Steve Ostro, who has just reported with several co-workers⁴ on all his observations during the past decade. The radars used are at Arecibo in Puerto Rico and Goldstone in California; the latter has a smaller range but is steerable, increasing the number of NEOs that may be accessed. For a large-scale search to give the orbits of NEOs to high accuracy on a single apparition (apparitions being several years apart), a Southern Hemisphere counterpart to these

Hole story

ASTRONOMERS believe they may have discovered a black hole in the distant galaxy M82 (J. T. Stocke *et al.* *Astr. J.* **102**, 1724–1728; 1991). A bright source of X-rays from the direction of M82 was discovered some time ago. Stocke and colleagues have now identified a massive star in the same position, and propose that the star and X-ray source are the two components of the brightest X-ray binary yet identified. In a typical X-ray binary, the X-rays are produced when material is dragged from a star onto a neighbouring neutron star. But excessive X-ray brightness would blow the material away, extinguishing the source. Given the luminosity of the M82 source, the authors think the mass of the accreting partner must be greater than 15 solar masses for its pull to overcome this limit — that is far too much for a neutron star, and so they conclude that the star they observe must be orbiting a black hole.

This little piggie. . .

THE search for new toys to play with is an important part of the life of the domestic piglet, but something that is often ignored in intensive pig-farming installations, say D. G. M. Wood-Gush and K. Vestergaard in *Animal Behaviour* (**42**, 599–606; 1991). In tests, piglets preferred to enter pens that bore the promise of some new object to examine, rather than one in which a familiar object had been concealed. An animal meeting a familiar object would turn tail and scamper towards the pen containing the unfamiliar one. Piglets, it seems, seek novelty for its own sake. This, propose the authors, is proof that young pigs display inquisitive exploration in which an animal acts to change its environment rather than merely responding to a stimulus.

Chemical fluff

MODERN mathematics meets classical chemistry in new calculations by D. Farin and D. Avnir, in *Angewandte Chemie (Int. Ed. Engl.)* **30**, 1379–1380; 1991), of the fractal dimension of a starburst polymer. The chemist's answer to the snowflake, starburst polymers grow radially from a central nucleus, each 'finger' of the polymer branching on successive generations or layers. In an extensive review last year in the same journal, D. A. Tomalia *et al.* showed that the surface area of the 'PAMAM' starburst polymer — both external and around voids in the interior — depends on the size, effectively, of the solvent molecules with which it is in contact. With the smallest molecules, indeed, there is more area on the inside than on the outside. Farin and Avnir use this simple property of the surface to calculate its fractal dimension, and find it to be 2.41.

radars would be highly desirable.

One thing that is not known with any certitude is the actual population of these smaller (10–100-metre) bodies, as until now they have been unobservable. Where do small asteroids stop dominating the flux, and fragments of comets take over in significance? The IAU resolution specifically defined NEOs as "minor planets [asteroids], comets, and fragments thereof", and we are largely ignorant of the distribution of sizes between those observed as bright meteors or fireballs in the atmosphere (the statistics are very poor for objects more than 1 metre across) and those viewed by telescope; few NEOs smaller than a few hundred metres across have been found. The best sample in the near future will come from the Spacewatch project, and the initial data indicate that there are rather more small NEOs than would be expected from a downward extrapolation from the larger bodies⁵.

The splitting of comets to form fragments is a phenomenon often observed, most recently in the case of the periodic comet Chernykh (1991o) in mid-September⁶. One of the most famous break-ups is that of comet Biela, two parts of which were observed in its apparition of 1846, and again in 1852, but neither since. Poulat Babadzhanyov and colleagues recently showed⁷ that comet Biela would have passed through the dense meteoroid cluster behind comet Tempel–Tuttle (the originator of the recurrent Leonid meteor storm observed every 33 years) in 1832, and that the 14 years that followed before the components were seen in 1846 would be appropriate to explain the physical separation of 250,000 km measured between those two fragments. It is a shame to spoil a good story, but Babadzhanyov *et al.* find from their frequency and energy that meteoroidal impacts are unlikely to have caused the splitting, and the significance of the calculations, they write, are "left for the reader to ponder over". The cause of cometary splittings, and related unexplained behaviour like the outburst of comet Halley earlier this year (see Paul Weissman's News and Views article⁸) remains an outstanding problem.

In this context, another significant recent discovery is asteroid 1991 RC, found by Robert H. McNaught as part of our own programme, AANEAS (the Anglo–Australian Near-Earth Asteroid Survey). This asteroid has an orbit whose size, shape and orientation are almost identical to those of 1566 Icarus⁹, which was discovered in 1949 and has an orbit quite unlike any other. It is plausible that 1991 RC and Icarus are twins, having resulted from the disintegration of a larger body. Coming hard on the heels of the recognition that Icarus has

an orbit perturbed by non-gravitational forces¹⁰ (due, presumably, to comet-like outgassing), this adds weight to the idea that asteroidal or cometary break-ups are commonplace, with substantial coherent complexes being formed which disperse over periods of 10⁴–10⁵ years. Considering that only one in a hundred NEOs larger than 0.5 km have yet been spotted, it seems likely that Icarus has other siblings awaiting discovery.

If NEO disintegrations result in the formation of complexes, do impacts on the planets occur randomly or in groups? In an open letter to the Morrison committee, Victor Clube has warned of the pitfalls of assuming that the hazard is dominated by the occasional (once per 100,000 years) stochastic impact by a 1-km or larger body: this would cause a global catastrophe and wipe out about half of the human race.

Contrary to this is the view of Clube's associates (myself included), which admits the existence of such events but believes the much more frequent low-level airbursts by 50–500-metre objects to be the principal risk to mankind. These are visualized as being largely confined to the coherent complexes formed in the hierarchical disintegration of large comets, with multiple impacts occurring in episodes lasting for a century or so when the orbital precession of each complex brings a node (ecliptic crossing) to the Earth's orbital radius; there is then a hiatus of many centuries until the next episode. Clube believes these are shown in the historical record¹¹. This might be termed the 'coherent catastrophism' school of thought, as opposed to the 'stochastic catastrophism' school. The idea is that the potential 1-km-plus impactors make up not a random, dispersed population, but must include recent acquisitions of such youth as to be associated with many smaller, shorter-lived bodies as well. The members of such complexes make up in numbers what they lack in individual size. The veracity or otherwise of this concept awaits confirmation as the NEO count is built up over the next few years.

Not least for its headline value, the hazard presented by NEOs has been foremost in astronomers' considerations. But Scotti *et al.* identify two other reasons for studying them: they are suitable candidates for space missions, often

1. Scotti, J. V., Rabinowitz, D. L. & Marsden, B. G. *Nature* **354**, 287–289 (1991).
2. Green, D. W. E. *IAU Circ.* 5367 (1991).
3. Chapman, C. R. & Morrison, D. *Cosmic Catastrophes* (Plenum, New York, 1989).
4. Ostro, S. J. *et al.* *Astr. J.* **102**, 1490–1502 (1991).
5. Rabinowitz, D. *Astr. J.* **101**, 1518–1529 (1991).
6. Luu, J. & Jewitt, D. *IAU Circular* 5347 (1991).
7. Babadzhanyov, P. B., Wu, Z., Williams, I. P. & Hughes, D. W. *Mon. Not. R. Astr. Soc.* **253**, 69–74 (1991).
8. Weissman, P. *Nature* **353**, 793–794 (1991).
9. Marsden, B. G. *IAU Circular* 5350 (1991).
10. Yeomans, D. K. *Astr. J.* **101**, 1920–1928 (1991).
11. Clube, S. V. M. & Napier, W. M. *The Cosmic Winter* (Basil Blackwell, Oxford, 1990).

being more accessible than the Moon; and their physical properties are of interest, not only for 'pure' motivations such as understanding their origin and evolution, but because they are a likely source of raw materials for major space construction in the next century. With regard to accessibility (as determined by the impulse needed for a rocket to reach it), the latest Spacewatch discovery, 1991

VG, is of especial interest. Also estimated to be 5–10 metres across, it will pass us by at a distance of 450,000 km (just further than the Moon) on 5 December, on an orbit that is strikingly Earth-like. □

Duncan Steel is at the Anglo-Australian Observatory, Private Bag, Coonabarabran, New South Wales 2357, Australia.

NUCLEOSYNTHESIS

Beryllium and the Big Bang

B. E. J. Pagel

HYDROGEN, helium and lithium should all have been created in the fireball that followed the Big Bang, but not beryllium, according to standard models — which explains the stir created by two groups^{1,2} who have detected beryllium in a star made of almost purely primordial material.

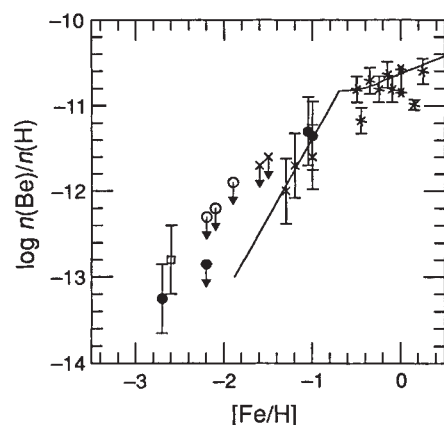
According to the standard model of Big Bang nucleosynthesis (SBBN)³, the early Universe expanded and cooled homogeneously from very high densities and temperatures, with radiation and relativistic particles (positrons, electrons, neutrinos and antineutrinos) remaining in thermal equilibrium for the first second or so. Thereafter, weak-interaction decoupling led to a virtual freezing of the neutron/proton ratio at about 15 per cent and positron–electron annihilation established the ratio of baryons (protons and neutrons) to photons which has remained unchanged to the present day. This ratio directly relates the measured temperature of the 2.75-K microwave background to the average density of ordinary matter.

Nuclear reactions set in about 100 seconds after the Big Bang, when the temperature became low enough to allow protons and neutrons to bind and form deuterons. The full theory gives a good fit to what has been deduced from astrophysical and cosmochemical measurements about primordial abundances (relative to hydrogen) of the light elements D, ³He, ⁴He and ⁷Li, provided that the baryon/photon ratio is $3\text{--}4 \times 10^{-10}$. Furthermore, the fit requires there to be no more than three types of light neutrino, a result strongly supported by measurements at CERN's Large Electron–Positron (LEP) collider. All other nuclear species are believed to result (along with additional He and ⁷Li) from nuclear reactions that took place long afterwards either in highly evolved stars (such as supernovae) or in the interstellar medium.

The composition of the medium is reflected, in turn, in that of the atmospheres of low-mass stars born from the medium at different times and still

observed today in, essentially, their original state. Thus, the oldest (unevolved) stars have virtually no carbon and heavier elements (referred to briefly, if inaccurately, as metals), and younger stars such as the Sun have maximal abundances.

The trouble is that the cosmic density of material predicted by the SBBN model is rather less than cosmologists would like. The standard against which density



The beryllium/hydrogen abundance ratio as a function of iron/hydrogen ratio for a sample of stars studied by Ryan *et al.*², Gilmore *et al.*¹ and others. The two left-hand points are for HD140283, a nearly primordial subdwarf. The solid line represents a standard model of beryllium evolution through normal galactic nucleosynthesis¹³. Is the difference attributable to a cosmological source of beryllium? (From ref. 2.)

can be compared is the critical closure density, which would allow the Universe to expand for ever, but only just, and would allow the geometry of the Universe to be 'flat'. This is the density preferred by cosmologists working with the inflationary picture of the Big Bang. Direct observations suggest that the ratio of the real density to the standard (Ω) could fall anywhere between about 0.1 and 1. But the SBBN calculations give Ω_B (the fraction of the closure density supplied by baryons) as no more than 0.06, and perhaps less, depending on the rate of expansion of the Universe (the Hubble constant). The difference

could be made up by hypothetical non-baryonic dark matter, for which there is no empirical evidence.

One way out of this impasse that has aroused much interest is that the baryon density could have been inhomogeneous during nucleosynthesis as a result of the quark–hadron phase transition which had taken place within a few tens of microseconds after the Big Bang (see Fukugita and Hogan's recent News and Views⁴). It had been thought that this could be a first-order transition (like that from steam to water) giving rise to supercooled bubbles of relativistic quark–gluon plasma that later resulted in fluctuations, both in baryon density⁵ and in the neutron/proton ratio, with potentially drastic effects on primordial nucleosynthesis⁶. From this, it was suggested that Ω_B could be 1 after all⁷ and that elements heavier than ⁷Li, even very heavy elements, could be produced in significant quantities in the neutron-rich regions⁸. Unfortunately for those enthusiastic for baryonic closure density, however, all versions of inhomogeneous Big Bang nucleosynthesis (IBBN) predict greater helium production than is compatible with observations if the averaged baryon density is more than a factor of 2 above the SBBN limit^{9,10}; this still falls short of $\Omega_B=1$ by a large and unbridgeable margin.

Nevertheless, if IBBN had actually occurred, it could have led to significant primordial abundance of nuclides not perceptibly produced by the standard mechanism. In particular certain versions of IBBN theory predicted that the light nuclide ⁹Be could have been produced in amounts that would be detectable in extremely metal-poor stars, even at low baryon densities¹¹.

The abundances of the light nuclides ⁹Be, ¹⁰B and ¹¹B in the Solar System can be pretty well accounted for in terms of spallation reactions between cosmic rays and interstellar C,N,O (ref. 12). But have their abundances evolved more or less in step with those of typical products of stellar thermonuclear reactions such as oxygen and iron? The answer depends on conditions of cosmic-ray acceleration and propagation in the earliest phase of the Galaxy, about which we know nothing. But if those conditions were not too different from those of today, then the production rate of beryllium and boron isotopes should have been lower in proportion to the lower interstellar abundances of C,N,O then prevailing. Consequently, their abundances in stars of low metallicity should be down by an additional factor comparable to the deficiency factors of C,N,O themselves¹³.

Observations of stellar beryllium and boron are quite difficult. The Be II doublet at 3,130 Å is in a crowded region of the ground-based ultraviolet

spectrum where the radiation flux from relevant stars is low; and B 1 at 2,496 Å can be observed only from space. However, the coming of the Hubble Space Telescope and of efficient echelle spectrographs and detectors for large ground-based telescopes are leading to a growth of new data in this field.

The oldest stars, with extreme metal deficiency (factors of 10 to some thousands below solar), belong to population II in the halo of our Galaxy and are rare, so that few examples are close and bright enough for high-resolution spectra to reveal beryllium and boron. One of the brightest is the subdwarf star HD140283, whose iron and oxygen abundances are, respectively, about 500 and 150 times below the solar values. The first attempt to detect beryllium in this star led to ambiguous results¹⁴, but the new data^{1,2} amount to convincing — if not totally indisputable — detections (see figure).

The implications are quite dramatic, as there is clearly more beryllium there than one would expect from cosmic-ray spallation (unless the oxygen abundance is higher as well). Are we seeing a cosmological floor of beryllium, analogous to that for helium and lithium? Probably not, as the whole idea of a first-order transition has been almost ruled out on theoretical grounds⁴. Boron has also been detected in HD140283 by the Hubble Telescope¹⁵. Conventional spallation calculations¹² predict a B/Be ratio of 10–30, whereas IBBN predicts less than 1. The observed ratio is apparently about 10.

My own bet is that some third process besides IBBN and spallation, for example photoerosion¹⁶, will turn out to be important in this story. But the question will not be decided until beryllium and boron abundances have been determined in other stars, especially in those with metallicities lower than HD140283. □

B. E. J. Pagel is at NORDITA, Blegdamsvej 17, DK-2100 Copenhagen O, Denmark.

1. Gilmore, G., Edvardsson, B. & Nissen, P. E. *Astrophys. J.* **378**, 17–21 (1991).
2. Ryan, S. G. et al. *Astrophys. J.* (in the press).
3. Yang, J. et al. *Astrophys. J.* **281**, 493–511 (1984).
4. Fukugita, M. & Hogan, C. J. *Nature* **354**, 17–18 (1991).
5. Witten, E. *Phys. Rev. D* **30**, 272–285 (1984).
6. Applegate, J. H. & Hogan, C. J. *Phys. Rev. D* **31**, 3037–3045 (1985).
7. Alcock, C., Fuller, G. M. & Mathews, G. J. *Astrophys. J.* **320**, 439–447 (1987).
8. Applegate, J. H., Hogan, C. J. & Scherrer, R. J. *Astrophys. J.* **329**, 572–579 (1988).
9. Kurki-Suonio, H. et al. *Astrophys. J.* **353**, 406–410 (1990).
10. Pagel, B. E. J. *Phys. Scripta* **T36**, 7–15 (1991).
11. Kajino, T. & Boyd, R. N. *Astrophys. J.* **359**, 267–276 (1990).
12. Walker, T. P., Mathews, G. J. & Viola, V. E. *Astrophys. J.* **299**, 745–751 (1985).
13. Vangioni-Flam, E., Casse, M., Audouze, J. & Obereto, Y. *Astrophys. J.* **364**, 568–572 (1990).
14. Ryan, S. G. et al. *Astrophys. J.* **348**, L57–L60 (1990).
15. Lambert, D. L. at 21st Int. Astr. Union Gen. Assembly, Buenos Aires (1991).
16. Boyd, R. N. & Fencil, H. S. *Astrophys. J.* **373**, 84–87 (1991).

The brave vertebrate venture

Henry Gee

THE transformation of fish into land vertebrates is pushed back another 10 million years this week, but the line between the two is increasingly hard to draw. On page 298 of this issue¹, Ahlberg redescribes what were once thought of as fish remains as those of tetrapods, helping to fill one of the most intriguing gaps in evolution — the transition from water to land.

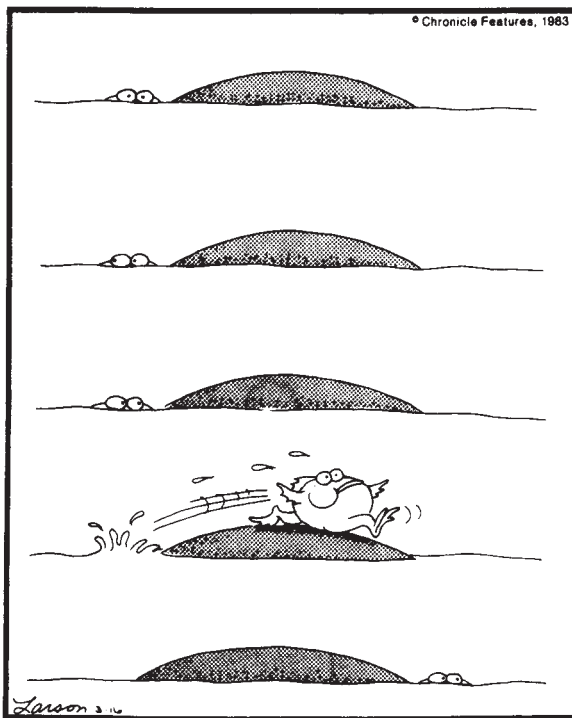
Until recently, cartoons have had more to say about the brave vertebrate

temporary *Acanthostega*^{4,5} has already started to close that gap, and Ahlberg now takes the process a little further.

The new material is a potpourri of bones from a little-known Upper Frasnian (Upper Devonian, about 370 million years old) exposure in Scotland. Jaw fragments once thought to represent fish have a characteristically tetrapod pattern of ornament, and features such as the closure of the meckelian canal in the lower jaw, and the existence of a humerus,

are strongly suggestive of tetrapod affinity. But the clincher is a tibia bearing articular facets for ankle bones — a *sine qua non* for feet, and thus tetrapods.

These results fit into a larger scheme in which the closest fossil relatives of tetrapods are thought to be an obscure group of fishes called the panderichthyids. Only two species are well known, *Panderichthys rhombolepis* from the Upper Frasnian of Latvia and the similar *Elpistostege watsoni* from the Upper Frasnian of Canada (which was, ironically, first described as a tetrapod⁶). These forms are definitely fish, but their precise phylogenetic position is unclear. Ahlberg regards them as closer to tetrapods than any other group, but Panchen and Smithson⁷ place them firmly within the osteolepid



Another great moment in evolution

venture onto land than have fossils. "I will be the first on land!", declares a fish floundering strandwards in a Larry Gonick drawing², dismissive of its companion's observation that the bugs had got there first³. "They won't get the credit" comes the reply: "my descendants will write the book". At the other extreme is Gary Larson's Far Side cartoon, "Another great moment in evolution", shown above.

Such was the gulf in the fossil record between fossil lobe-finned fish and the erstwhile earliest-known tetrapod *Ichthyostega* (often reconstructed to be as unfishlike as one could imagine save for its fishlike tail) that Gonick is entitled to poke fun at anyone brave enough to spin a yarn strong enough to cross it. But palaeontology is beginning to catch up. Evidence for fishlike internal gills and ventilation in *Ichthyostega's* Upper Famennian (uppermost Upper Devonian, about 365 million years old) con-

lobe-finned fishes, making them no more closely related to tetrapods than, say, the well-known osteolepid *Eusthenopteron*, often cast as the sort of fish one would expect to have as an ancestor.

But there are other, shadowy forms known only from a few scraps. Ahlberg's Scottish tetrapod may be congeneric with the Latvian *Obruchevichthys gracilis*, usually thought of as panderichthyid. And this summer, Ahlberg and Oleg Lebedev unearthed new material of '*Panderichthys*' *bystrowi*, demonstrating that this poorly characterized form can now be numbered among the earliest tetrapods along with *Obruchevichthys* and the Scottish material.

Whatever one's views about the phylogenetic placement of the panderichthyids, it is clear that these or similar fishes had begun to sprout legs as early as the Frasnian. But, to paraphrase Thornton Wilder's schoolgirl pianists, in his play *Our Town*, they weren't happy

Gary Larson/Chronicle Features, San Francisco

ISOTOPE DATING

about it. A few million years later, early tetrapods were still firmly tied to the water. Enough work on the postcranial material of *Acanthostega* has now been done for M. I. Coates to present a full reconstruction at a meeting in London last month, in which the aquatic adaptations of the skull, limbs and shoulder girdle were amply demonstrated.

Fish differ from tetrapods in that their shoulder girdles are firmly attached to the back of the skull by a sequence of dermal bones that are reduced or lost in tetrapods. *Acanthostega* retains a fishlike cleithrum and anocleithrum (bones of the shoulder girdle) that were probably involved in supporting the gill chamber⁵, and which are similar to those in the living lungfish *Neoceratodus*. The forelimb, which like those of all known early tetrapods had supernumerary digits^{8,9}, looked like a flipper and would have been capable of little more than flapping up and down.

The fish pelvis is typically a single plate that is not attached to the vertebral column, contrasting with the tetrapod arrangement of ilium, ischium and pubis, connected to the vertebral column through the ilium. The *Acanthostega* pelvis appears to be a single plate but its connection or otherwise to the vertebral column remains obscure. The hindlimbs — now thought to have eight digits apiece, like the forelimbs — jutted out at the sides in the manner of hydroplanes. *Ichthyostega*'s forelimbs were stouter than those of *Acanthostega* but the hindlimbs were similarly flipper-like.

These fossil glimpses are tantalizing enough, but it seems clear that the animals we proudly call our ancestors were thoroughly aquatic, breaking the surface only in the direst of circumstances. Certainly, the suite of adaptations needed by a fish out of water to avoid being asphyxiated, desiccated and crushed under its own weight would be enough to deter the hardiest adventurer. If Gonick pokes fun at what was once popularly thought to have been a dignified parade from water to land, Larson's irony points out that those of us, the descendants who write the book, are all too prone to gloss over the fact that permanent terrestrial residence might have been nothing more than an accident waiting to happen. □

Henry Gee is an assistant editor of *Nature*.

1. Ahlberg, P. E. *Nature* **354**, 298–301 (1991).
2. Gonick, L. *The Cartoon History of the Universe* (Rip Off Press, San Francisco, 1978).
3. Shear, W. A. *Nature* **351**, 283–289 (1991).
4. Clack, J. A. *Nature* **342**, 425–427 (1989).
5. Coates, M. I. & Clack, J. A. *Nature* **352**, 234 (1991).
6. Westoll, T. S. *Nature* **141**, 127–128 (1938).
7. Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341–438 (1987).
8. Lebedev, O. A. *Dokl. Akad. Nauk. SSSR* **278**, 1407–1473 (1984).
9. Coates, M. I. & Clack, J. A. *Nature* **347**, 66–69 (1990).

Quick shifts in plate motion

Bruce W. D. Yardley

GEOLOGICAL processes have a reputation for grinding exceedingly slowly. Even rapid events such as earthquakes or volcanic eruptions are part of a continuum of catastrophes affecting a part of the Earth's crust over a long time. Sometimes, however, as illustrated by Goldfarb *et al.* on page 296 of this issue¹, the patterns of plate motion do change and geological activity can begin remarkably quickly in previously stable areas.

Quite how abruptly such changes happen is difficult to resolve — in geological terms they can be almost instantaneous.

For example, the central volcanic zone of the North Island of New Zealand has developed in just the past 1 million years. But the geochronological precision necessary to determine how rapidly new geological phenomena have developed in the past, has not been attainable. The dating of gold-bearing quartz veins described by Goldfarb *et al.* provides a new insight into the effects of rapid changes in plate motions and their wider geological implications. The authors have performed argon-isotope dating of muscovites from gold-quartz veins

The great wooden well of Kückhoven

WHY should a prehistoric village have been sited some 3 km from the nearest watercourse, surely an uncomfortable distance to fetch and carry one of life's necessities? That puzzle has been solved by the discovery of what can lay fair claim to being the oldest surviving wooden structure in the world — a great well, dating to 5303 BC — at Kückhoven, south of Mönchengladbach, in northwest Germany. The remarkable state of preservation means that the construction technique used by its Neolithic builders, and even the tools used, can be assessed.

Since mid-1989, a team of archaeologists from the Landschaftsverband Rheinland, led by Jürgen Weiner, have been excavating a gravel-pit at Kückhoven, where the remains of a settlement of the Linearbandkeramik culture are under threat and will be destroyed by gravel extraction in a few years. Most Linearbandkeramik settlements were villages of agriculturalists and pastoralists, and post holes uncovered by the team show that the people lived in long timber houses. The sites of the period are normally located on good loess soils near streams or rivers — but not that of Kückhoven, which lay on an expanse of loess but seemed to be nowhere near a source of water.

Late last year the first fragments of wood were found, and subsequent excavation this summer located, at a depth of 7 m below the present ground surface, a shaft with vertical walls of 5 or 6 m, with abundant wooden fragments in its centre. A test-pit revealed a further 7.6 m of shaft below and, in the middle, two rectangular well-frames in wood. The outer one is older, and is 3 m square, while the inner, placed in its northwest corner, is 1.6 m square, and seems to have been built after the older one was damaged. Fragments of wickerwork suggest that the upper part of the shaft may have been lined with this material.

The 200 oak planks discovered so far are up to 15 cm thick and 50 cm wide. After being felled, the trees must have had their trunks split with the stone axes of the period. The planks bear clear traces of these stone tools, while other tools, probably chisels made of cattle bones, were used to cut rectangular mortises in them. In this way, the beams could be joined into the box-frame, the slots being caulked with moss. The remains of several tools have been recovered in the well, including two composite mattocks with wooden blades and hafts. These are the only such tools ever recovered from an Linearbandkeramik settlement. More finds can be expected when the shaft is completely emptied, because wells which fell into disuse — perhaps after drying up — were often used as rubbish tips.

Initial results from radiocarbon dating have been confirmed and made more precise by dendrochronological analysis of the tree-rings on the planks. It is now possible to say that the oak trees were felled in 5303 BC exactly, probably in the autumn or winter. The well was built in the following spring or summer. Its shaft was more than 15 m deep, the lowest 8 m being beautifully preserved.

The Kückhoven find is notable in two respects. First, it makes clear that the people of the time had mastered comparatively sophisticated carpentry techniques. Second, analysis of its timbers is making a crucial contribution to the dendrochronological curve for this period in Europe. The remarkable conditions of preservation, caused by a combination of soil chemistry and the dampness around the structure, have ensured that a Rhineland museum will eventually be able to display a unique and perfectly conserved example of Neolithic prowess in engineering.

Paul G. Bahn

Paul G. Bahn, 428 Anlaby Road, Hull HU3 6QP, UK, is a freelance writer on archaeology.

along a 200-km length of the Juneau gold belt of Alaska. The veins were formed at depths of more than 5 km (but probably less than 10 km) and at temperatures of around 300 °C, and as such are typical of the so-called mesothermal gold deposits which crosscut many of the world's metamorphic belts. Similar vein systems provoked gold rushes in the United States, formed the basis of the wealth of European cities such as Salzburg, and have provided the British royal family with their wedding rings. The fluids which form such deposits appear to be remarkably uniform. Typically they are low-salinity $\text{H}_2\text{O}-\text{CO}_2$ fluids containing CH_4 , N_2 and H_2S .

Although it has long been realized that deposits of this type probably form over relatively short periods of time, Goldfarb *et al.* have obtained ages precise to significantly better than 1 million years, and demonstrate for the first time the true brevity of the vein forming event. Ages along a 200-km vein system all lie between 55.0 and 56.1 million years. It is clear from both the geological context and the argon release profiles that these are true ages of formation and not the result of some subsequent resetting of the argon clock. At about this time in the early Eocene there was a significant change in the angle of convergence of the Kula Plate (part of the Pacific Ocean floor, now completely subducted) relative to the North American Plate, and Goldfarb *et al.* suggest that the veining event was triggered by the shift away from compressional tectonics. What we see here, however, is not a shift from one steady state to another, but a singular, brief event, triggered by the change in plate motion.

Quartz is not a particularly soluble substance, even at 300 °C; indeed the presence of CO_2 and other non-polar fluid species serves to suppress its solubility. The quartz veins probably formed from fluids that carried at most only a few hundred parts per million of silica in solution. The presence of elevated gold concentrations and the alteration of wall rocks rule out an origin for the vein quartz by local segregation from nearby rocks. These veins clearly represent major pathways of focused fluid flow through the crust. As such, the rapid formation of the gold-quartz veins acquires further significance because they may also throw new light on the contentious issue of whether fluid is present deep in the Earth's crust.

Some geophysicists² have argued that significant quantities of water are present deep in stable continental crust at the present time, and so account for the electrical conductivity of the deep crust. It has also been suggested that such saturated zones may give rise to seismic reflections. On the other hand, meta-

morphic petrologists point out that any water remaining in metamorphic rocks once they have begun to cool will be rapidly absorbed by retrograde reactions producing hydrous minerals³. The extensive preservation of anhydrous, high-temperature metamorphic minerals suggests that the rocks were dry throughout their post-metamorphic history. Goldfarb, with additional co-workers, has recently published isotope data on



A typical late metamorphic quartz vein, from Val d'Ayas, Italy, one of a suite that has been mined for gold. CO_2 in the fluids responsible for the vein altered the wall rocks to give brown-weathering carbonate minerals.

minerals from the Juneau veins⁴, which they believe provide clear evidence that the vein-forming fluids were indeed metamorphic waters from deep in the crust that had survived at depth by some kind of ponding process. In this issue, he and his colleagues propose that the veins formed when the changing regional stress pattern allowed them to break through an impermeable cap layer. Thus the veins form in a brief event as deep over-pressured fluid reservoirs are drained along new fractures.

Clearly this is an attractive and cogent hypothesis, but it is unlikely to sweep aside all doubts about the persistence of deep fluids. The principal alternative is that the vein fluids moved down from the surface. Unfortunately, testing between such radically different alternatives will remain difficult because at elevated temperatures fluids are continually exchanging material their hosts so that any geochemical memory of the nature of the fluid at source becomes blurred or lost. □

Bruce W. D. Yardley is in the Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.

1. Goldfarb, R. J., Snee, L. W., Miller, L. D. & Newberry, R. J. *Nature* **354**, 296–298 (1991).
2. Gough, D. I. *Nature* **323**, 143–144 (1986).
3. Yardley, B. W. D. *Geology* **9**, 405–408 (1981).
4. Goldfarb, R. J., Newberry, R. J., Pickthorn, W. J. & Gent, C. A. *Econ. Geol.* **86**, 66–80 (1991).

Broken dreams

TELEVISION, says Daedalus, is the most pernicious of inventions. It may appear to convey information, but in fact it conveys dreams. The unspoken, unchallenged, unrealizable ideals of whole societies are now defined by television. Worse, because television is extremely expensive to set up and run, the ideals it broadcasts are those of the rich. The effects on developed societies are bad enough. But global satellite television is now beginning to taunt the vast population of the underdeveloped world with dreams of riches that are unfeasible even to the industrial nations. The resulting mass envy and frustration must inevitably lead to some kind of global social disaster.

So Daedalus is looking for some way of sabotaging satellite television, preferably without drawing attention to the fact. Accordingly, he has developed a renewed interest in the Earth's threatened ozone layer. He advocates the replacement of halocarbon fluids in refrigerators and air-conditioners by the traditional refrigerant, ammonia. He is also reviving an old scheme of his to stabilize the ozone layer with direct injections of ammonia. In the stratosphere, the ammonia will be photo-oxidized to nitrogen oxides; these will combine with the crucial free radicals responsible for the chain reaction that destroys the ozone. As militant ecologists denounce the over-fertilization of the land by ammonia fertilizers, the chemical industry should be happy to find an environmentally benevolent outlet for millions of tons of surplus ammonia. It should eagerly support Daedalus's plan to inject the gas into the stratosphere by means of huge vortex-ring generators to squirt it upwards. Because ammonia is lighter than air it should rise anyway; deliberate emissions and releases from leaky air-conditioners and discarded refrigerators will all accumulate in the stratosphere.

Daedalus's hidden agenda is very simple. Ammonia has a wide range of strong absorption bands across the 10–18 GHz region which contains the satellite-TV uplink and broadcast frequencies. As his virtuous attempts to save the ozone layer develop, global television will be rapidly degraded. A satellite channel has no power to spare for attenuation losses, even with sensitive big-dish receivers. Relatively little atmospheric absorption should black it out completely. With luck it should not be at all obvious what is happening. To get any signal through at all, global broadcasters will be forced to narrow their bandwidth right down, until there is only enough space for simple radio. And radio really does transmit information, not dreams. David Jones

Activated C₇₀ and diamond

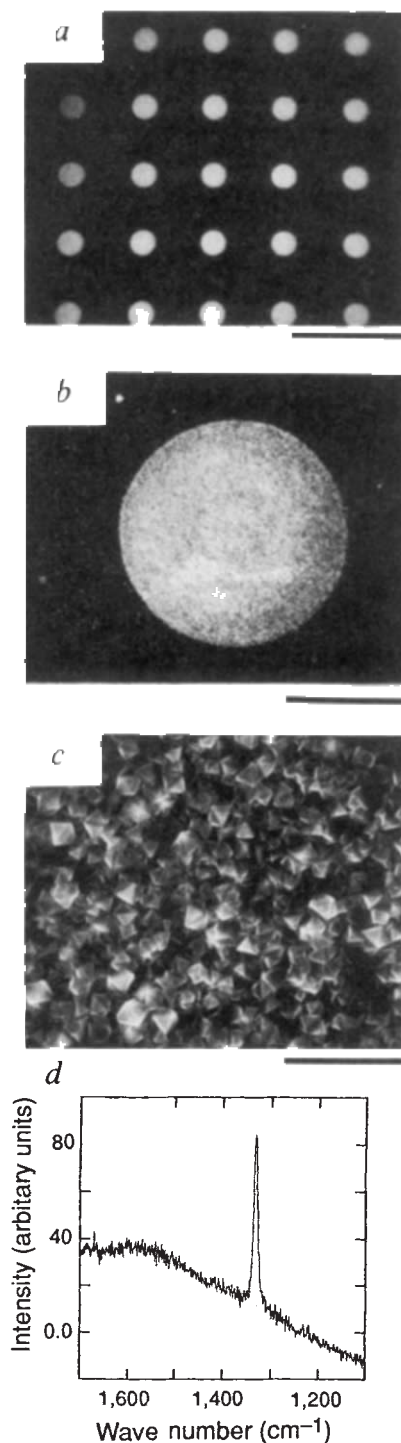
SIR — We have developed a novel method for nucleating high-density diamond crystallites by substituting thin, solid-film C₇₀ carbon clusters for diamond grit polish. Our method also allows for the one-step lithography of diamond growth in preselected areas.

Chemical vapour deposition (CVD) is a promising technique for growing diamond films on various substrate materials for protective coatings and the fabrication of active electronic and optical devices. To obtain high nucleation density of diamond crystallites on non-diamond substrates to form continuous films, a pretreatment of diamond grit polishing is necessary, which limits the full technological development of the CVD process.

In our experiments, 50–100-nm-thick films of pure C₇₀ have been thermally sublimated onto semiconductor-grade silicon wafers and other substrates (such as Mo, SiO₂). Diamond films are grown on these surfaces in a microwave (2.45-GHz) plasma reactor¹ with substrate temperature at 900°C and in a medium of 1% CH₄ in H₂ at about 100 torr total pressure. Unlike hydrocarbon molecules which are unstable in a hydrogen plasma, C₇₀ molecules are rather stable. Thus, we need to activate the cluster films for diamond nucleation, with a pretreatment of positive ion bombardment by biasing (200–300 V) the substrate with respect to the plasma. During the pretreatment (several minutes), the gas composition is 10% CH₄ in H₂ at 15 torr total pressure. Following this pretreatment diamond growth takes place. The figure (a) shows a scanning electron micrograph of diamond crystallites grown after activation on an array of 100-nm-thick C₇₀ dots on an Si substrate (formed by shadow masking) with a closer view of an individual dot in b. The figure also shows (c) the grain size (about 1 µm) and morphology of the resulting continuous (111)-oriented diamond film. The nucleation density in this film is comparable to that found on a diamond-grit polished Si substrate. Thus the amount of nucleation enhancement is close to 10 orders of magnitude relative to an untreated silicon surface. Other substrates, such as Mo and SiO₂, have also been used and show similar results. We have also used C₆₀ dots, but with less success (only a few orders of magnitude of enhancement). The Raman spectra of the diamond film (c) grown on activated C₇₀ are shown in d.

To assess the relative use of C₇₀ cluster films for diamond nucleation, we have used other forms of carbon such as diamond-like carbon, highly oriented pyrolytic graphite and polycrystalline

graphite. These forms enhance nucleation by only a few orders of magnitude under the same activation and growth conditions. In addition, we have found that hydrocarbon molecules, such as hydrocarbon-based oil and adamantane, are unstable in a hydrogen plasma, espe-



a, b, Diamond growth on 100-nm-thick C₇₀ dots (200 µm in diameter). c, Scanning electron micrograph of (111)-oriented diamond crystallites grown on a C₇₀ film. Scale bars: a, 0.75 mm; b, 100 µm; c, 10 µm. d, Raman spectra of the film in c.

cially above a temperature of 300°C.

From our experimental results we can speculate why C₇₀ is a unique precursor for diamond nucleation in a plasma CVD environment. Compared to other precursors, C₇₀ is more stable in a hydrogen plasma at high temperatures. Unlike graphite, C₇₀ provides spatially a third dimension (within the shell) to allow easy conversion from *sp*² bonding to *sp*³ bonding which is required for diamond nucleation and growth. Along the short axis perimeter of C₇₀ there are five ensembles of four hexagons which can be restructured to produce chair-form cyclohexane linkages. These linkages are representative of the surface of the 111-plane of diamond from which diamond can grow².

R. MEILUNAS

R. P. H. CHANG

Department of Materials Science and Engineering,

SHENGZHONG LIU

MANFRED M. KAPPES

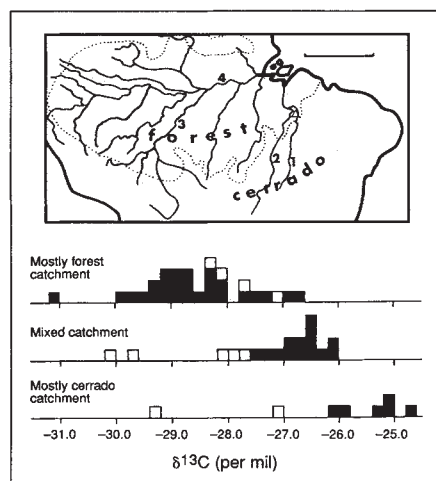
Department of Chemistry,
Northwestern University,
Evanston, Illinois 60208, USA

1. Meilunas, R. & Chang, R. P. H. in *Proc. 2nd ICAM Conf.* (eds Chang, R. P. H., Geis, M., Meyerson, B., Miller, D. & Ramesh, R.) 609–614 (Materials Research Society, Pittsburgh, 1990).
2. Tsuda, M., Nakajama, N. & Okawa, S. *J. Am. chem. Soc.* **108**, 5780–5784 (1986).

Carbon ratios in the Amazon

SIR — Despite the importance of the rainforests of the Amazon basin to greenhouse gas fluxes, terrestrial carbon storage and global biodiversity^{1,2}, there remains no consensus as to the nature of vegetation present in the region during the Last Glacial Maximum (LGM)³. Although there is broad agreement that the climate was cooler, opinions diverge on the question of glacial aridity in the region, and hence on the nature and extent of closed forest cover. The major barrier to resolving this problem has been that many proxy vegetation and climate records yield only fragmentary information which is poorly time constrained, and which pertains only to a localized region.

The carbon-isotope composition of organic matter in fluvial sedimentary sequences offers the possibility of investigating past vegetation changes on a basin-wide scale. The modern Amazon forests are almost exclusively dominated by plants which photosynthesize via the C₃ pathway, with δ¹³C values of about –28‰. In contrast, the more open ‘cer-rado’ vegetation (grassland/woodland) on the margins of the forest, contains a variable proportion of tropical grasses which photosynthesize via the C₄ pathway, yielding δ¹³C values of about



Carbon isotope results and study area. 1, Tocantins; 2, Araguaia; 3, Madeira; 4, Amazon rivers. Scale bar, 1,000 km. Dotted line, present extent of forest; triangle, general area in which sampling was undertaken for this study. The $\delta^{13}\text{C}$ results are reported in per mil, relative to the PDB standard. Open boxes, dry season; closed boxes, wet season. Data for forested catchments from ref. 5. Data for cerrado catchment (Tocantins) from this study. Data for mixed catchments from this study (Araguaia, and two unnamed streams south of Porto Franco-Pará) and ref. 5 (Madeira).

–12‰ (ref. 4).

The $\delta^{13}\text{C}$ values of particulate organic carbon (POC) in fluvial sediments will depend upon the proportion of C_3 - and C_4 -derived carbon in the vegetation and soils of the tributary catchments (algae are characterized by low $\delta^{13}\text{C}$ values, but do not contribute greatly to the carbon budget of rivers in the Amazon basin⁵). Therefore, if the Amazon forest contracted to isolated 'refugia' during the LGM, and was largely replaced by cerrado-style vegetation, a shift to more positive $\delta^{13}\text{C}$ values should be observed in sediments of that age.

To test this hypothesis, suspended sediment and bottom sediment samples were collected from the Tocantins/Araguaia river system, on the eastern margin of the basin. This is the only major tributary of the Amazon with a catchment that is not predominantly forested, and one of the few tributaries for which there is not already a substantial $\delta^{13}\text{C}$ database. The samples were pretreated and analysed according to the procedures outlined by Hedges *et al.*⁵ to make the results directly comparable with the data from the other major tributaries detailed in that study.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

The figure shows that there is a difference of 2–4‰ between the $\delta^{13}\text{C}$ values of POC from rivers draining forest versus cerrado catchments during the wet season, but no difference during the dry season. The lack of difference in the dry season results from the fact that most rivers in the cerrado region are fringed by a C_3 -dominated strip of forest, which provides the bulk of the carbon exported during the dry season (–27 to –30‰). During the wet season, high runoff rates transport C_4 carbon to the river from more remote areas leading to $\delta^{13}\text{C}$ values as high as –24.6‰.

The $\delta^{13}\text{C}$ value of the POC in rivers draining forested catchments is the same as that of the forest vegetation during both the wet and dry seasons. However, in rivers draining cerrado catchments, the $\delta^{13}\text{C}$ values of POC are consistently lower than in the bulk vegetation (cerrado soils have $\delta^{13}\text{C}$ values between –15 and –25‰). This is due to the above-mentioned bias towards C_3 carbon derived from close to the river, preferential metabolism of C_4 carbon by microorganisms in the river, and a small contribution from low- $\delta^{13}\text{C}$ algal biomass.

While the $\delta^{13}\text{C}$ value of POC from rivers draining cerrado catchments does not accurately reflect the bulk $\delta^{13}\text{C}$ value of biomass in the catchment, the cerrado $\delta^{13}\text{C}$ signature can nonetheless be readily distinguished from the $\delta^{13}\text{C}$ signature of POC in rivers draining forested catchments. Therefore, the $\delta^{13}\text{C}$ value of organic matter in ancient fluvial sediments can potentially provide a basin-wide history of forest growth/collapse during the Quaternary both in the Amazon basin and in other tropical regions. Interpretation of $\delta^{13}\text{C}$ trends in ancient sediments also need to take into account changes in the concentration and $\delta^{13}\text{C}$ value of atmospheric CO_2 (refs 1, 6) changes in the proportion of previously respired CO_2 re-utilized by forest vegetation⁷, and the possibility of diagenetic modification of the $\delta^{13}\text{C}$ value of POC after burial.

MICHAEL I. BIRD

ALLAN R. CHIVAS

Research School of Earth Sciences,
The Australian National University,
PO Box 4, Canberra, ACT 2601,
Australia

WILLIAM S. FYFE

Geology Department,
University of Western Ontario,
London, Ontario N6A 5B7,
Canada

1. Crowley, T. J. *Nature* **352**, 575 (1991).
2. Soili, H. (ed.) *The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and its Basin* (Junk, Boston, 1984).
3. Colinvaux, P. A. *Nature* **340**, 188 (1989).
4. Smith, B. N. & Epstein, S. *Plant Physiol.* **47**, 380 (1971).
5. Hedges, J. I. *et al.* *Limnol. Oceanogr.* **31**, 717 (1986).
6. Freidli, H. *et al.* *Geophys. Res. Lett.* **11**, 1145 (1984).
7. Van der Merwe, N. J. & Medina, E. *Geochim. cosmochim. Acta* **53**, 1091 (1989).

Carbon suboxide in Halley

SIR — Crovisier *et al.*¹ conclude from an examination of the VEGA 1 IKS infrared spectra² that carbon suboxide (C_3O_2) is not present in the comet Halley ices at the abundance level suggested by us³. They show the measured spectrum near 4 μm and note the absence of the strong C_3O_2 line emission, relative to the CO_2 emission, that should be present if the C_3O_2 production rate was comparable³ to that of CO_2 . Key assumptions in this analysis are that C_3O_2 is released from the comet nucleus and has a scale length similar to the CO_2 scale length.

Our discussion³, however, is based on a coma grain source for C_3O_2 peaking at a distance from the nucleus (R) of $\sim 8 - 10 \times 10^3$ km. For a radial velocity of 1 km s^{-1} , the scale length of C_3O_2 is only $\sim 2 \times 10^3$ km, whereas the scale length of CO_2 is $\sim 4 \times 10^5$ km³⁻⁶. The concentration, c , of a species flowing radially outward from the surface of a sphere (with radius R_0) and decaying exponentially with a scale length H is⁷

$$c(R) = c(R_0) (R/R_0)^2 e^{-(R-R_0)/H}$$

If CO_2 is released from the comet nucleus (mean $R_0 \sim 7$ km) and C_3O_2 enters the gas-phase at $R_0 \sim 10^4$ km with equal production rates, integration of the equation yields a maximum column density along the line of sight to the nucleus for CO_2 that is 10^4 times the calculated column density of C_3O_2 . Even with the C_3O_2 fluorescence factor three times larger than CO_2 , the C_3O_2 infrared emission should be much weaker at 4 μm than the CO_2 emission. This is consistent with the observational data shown by Crovisier *et al.*

Our brief analysis suggests that C_3O_2 can have a large coma production rate, relative to water, but still be hard to detect by remote sensing because of the large area over which it enters the coma gas-phase. A more detailed investigation taking into account the viewing geometry of the IKS experiment is needed to assess the expected infrared emission from C_3O_2 arising from a distributed coma dust source.

MARK ALLEN

Jet Propulsion Laboratory,
California Institute of Technology,
Pasadena, California 91109, USA

1. Crovisier, J., Encrenaz, T. & Combes, M. *Nature* **353**, 610 (1991).
2. Combes, M. *et al.* *Nature* **321**, 266–268 (1986).
3. Huntress, W. T., Jr, Allen, M. & Delitsky, M. *Nature* **352**, 316–318 (1991).
4. Nakata, R. S., Watanabe, K. & Matsunaga, F. M. *Sci. Light* **14**, 54–71 (1965).
5. Hitchcock, A. P. & Brion, C. E. *Chem. Phys.* **45**, 461–478 (1980).
6. Lewis, B. R. & Carver, J. H. *J. Quant. Spectrosc. Radiat. Transf.* **30**, 297–309 (1983).
7. Haser, L. *Bull. Acad. R. Belg., 5^e série* **43**, 740–750 (1957).

Long-lived transitional states of the geomagnetic field and the two dynamo families

Kenneth A. Hoffman

Physics Department, California Polytechnic State University, San Luis Obispo, California 93407, USA

Palaeomagnetic data from lavas produced at the Hawaii and Society Islands hotspots suggest that one or more specific, long-lived field configurations recur during successive polarity reversals spanning perhaps several million years. The apparent quasi-stationarity of these states and the relative placement of the two recording sites offer a rare opportunity to explore the fundamental geometrical properties of transitional fields.

OUR understanding of the process by which the Earth's magnetic field reverses comes largely from the palaeomagnetic record (see refs 1–5 for recent reviews and articles). Each transition record, in principle, provides information about vector field changes experienced at the recording site. Yet the analysis of global field behaviour during reversals is plagued by limitations, especially those imposed by differing palaeomagnetic signatures and degrees of record fidelity, insufficient global coverage of particular events, and when contemporaneous records do exist, our often wholly inadequate knowledge regarding synchronicity.

Notwithstanding these limitations, from the time the first palaeomagnetic recordings became available, attempts were made to interpret the gross structure of transitional fields. A debate rapidly arose over whether simple dipole configurations, or alternatively, more complex non-dipole morphologies, dominate the reversal process^{6,7}. Although non-dipolar interpretations for transitional fields have been fashionable for more than a decade^{1–5}, the observation that reversal paths of the virtual geomagnetic pole (VGP) run preferentially along particular bands of longitude^{8–10} has recently refuelled the controversy¹¹ with discussion of dipolar solutions. Yet it is a fundamental property of the geomagnetic dipole that, when compared with any non-dipole component, its relative strength will be progressively enhanced with distance from the outer core. Simply stated, the further one is from the field source, the more the dipole term will dominate. Moreover, theory suggests the existence of two dynamo families¹² in which there is no simple distinction between dipole and non-dipole geometries; each family contains both^{12,13}. Hence, the question of dipole versus non-dipole configurations may have limited significance to the issue of the morphology of reversing fields. Here I ask a different question, first posed by Merrill and McFadden¹⁴: whether either dynamo family typically dominates the geomagnetic field when in transition. The answer to this question may aid our understanding of dynamo sources and behaviours relevant to the reversal mechanism¹⁵.

I present here evidence from Plio-Pleistocene lavas from two Pacific hotspots (refs 16–19, and this study) that the process of polarity reversal involves the development of specific quasi-stationary transitional field configurations which recur over timespans involving several events. Because the observation sites under consideration are disposed almost symmetrically about the Equator, the recording of such long-lived states makes possible the consideration of palaeofield symmetry with respect to the Equator. Equatorial symmetry and antisymmetry distinguish fields corresponding to the two dynamo 'families'^{12,13}; here I develop tests for family dominance in terms of

palaeodirections as well as their corresponding VGPs. For example, a case can be made that the recorded long-lived transitional field states are dominated by members of the dynamo family having symmetry about the Equator, and thus occur when members of the dynamo family having antisymmetry about the Equator—in particular, the axial dipole—essentially vanish for a time. For this model, the degree to which transitional states are dipolar depends on the strength of the equatorial dipole term at the Earth's surface relative to that of the other (non-dipole) members of the family present during the reversal.

New transitional data from Molokai, Hawaii

At Hakaano Peninsula [21° N, 203° E] on the island of Molokai, in the Hawaiian chain, a sequence of basalts was extruded during the normal to reverse (N–R) polarity change at the end of the Olduvai subchron (1.7 Myr ago). Stepwise demagnetization in alternating fields was applied to specimens taken from cores drilled at this site to identify the primary component of magnetization. The resulting palaeodirections, determined through principal component analysis²⁰, are plotted in rotated directional (D' , I')-space²¹ (Fig. 1; D and I indicate declination and inclination). As can be seen, few flows recorded directions deviating by more than 30° from that of the axial dipole. The intermediate directions which are found for the most part define a relatively tight cluster about $D = 134.7^\circ$, $I = -63.5^\circ$; the corresponding VGPs are located in the Atlantic off the coast of South America (Fig. 1).

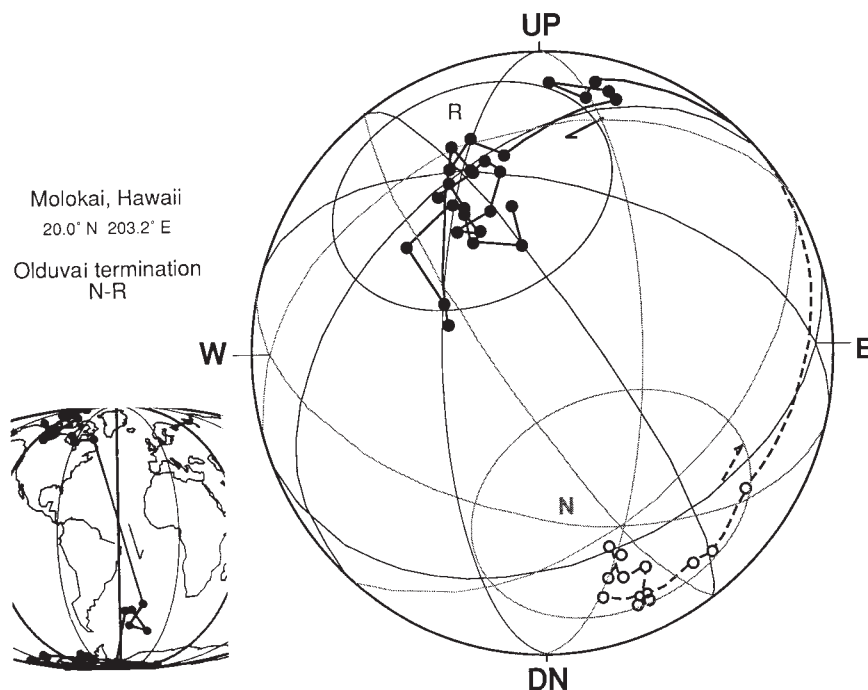
Given the paucity of transitional flows and the fact that the variability in extrusion rate for these lavas is not known, one might conclude that the recorded directional group has no special significance to the process of field reversal. This argument notwithstanding, an intriguing picture emerges when these particular data from Molokai are compared with those associated with other Plio-Pleistocene reversals obtained from lavas erupted at the Hawaii hotspot.

Other records from the Hawaiian hotspot

One such case is the detailed record of the younger (0.73 Myr) Matuyama-Brunhes (R–N) polarity transition obtained from lavas on Maui, Hawaii^{17,22}. Specifically, this recording begins with a tight cluster of palaeodirections defined by no fewer than 22 successive lavas¹⁷. The average direction associated with this cluster is $D = 105.1^\circ$, $I = -29.7^\circ$, indicating a VGP within South America, relatively close to the position found from the Molokai record.

Another reported transition record obtained from lavas extruded at the site of the Hawaii hotspot, now exposed on the island of Kauai^{16,23}, is of Pliocene age. This R–N transition is defined almost exclusively by a tight clustering of more than 20 flow directions having average orientation $D = 154.9^\circ$, $I = -22.0^\circ$. The corresponding VGP for this case is found south of the tip of South America. Within the sequence of lavas that record this cluster there exists a flow possessing a direction consistent with full normal polarity. The most reasonable interpretation is that the Kauai lavas record a rebound^{6,16,24} of the geomagnetic field. As there is an almost complete lack of other transitional directions, the duration associated with the recording of the directional cluster (both before and after an apparently unsuccessful attempt by the field to reverse) is probably a considerable fraction of the total time taken for the complete

FIG. 1 Palaeodirections corresponding to the Olduvai-Matuyama (N-R) transition recorded on Molokai, Hawaii, plotted in (D' , I')-space²¹. The vector data are projected sequentially from the centre of a sphere and are geometrically displayed as though observed in their actual geographic surroundings. The 'pseudo-poles' correspond to the normal (N) and reverse (R) axial dipole field directions for the site. The 'pseudo-Equator' is the great circle containing all possible directions 90° (midway) from the axial dipole directions and the small circles (shown for scale) correspond to directions making a 30° angle from that of the nearest axial dipole. Inset: path of the corresponding VGP (that is, the sequence of geomagnetic poles that would occur on the Earth if the recorded directions were associated with purely dipole palaeomagnetic fields).



reversal. Such a conclusion is compatible with other reported transitional records²⁵⁻²⁷.

The average VGP associated with each of the three directional groups described above, which span reversals from ~4.5 Myr ago (S. Bogue, personal communication) to the start of the Brunhes normal polarity chron, is plotted in Fig. 2. Although one cannot rule out the possibility that their similarity is coincidental, these reversal data are consistent with the hypothesis that the process of field reversal involves quasi-invariant features that recur, perhaps, over several transitions^{8,28,29} and, moreover, that one such feature is the occurrence of one or more stationary intermediate field configurations which may dominate over a significant fraction of the reversal duration²⁵⁻²⁷.

Record from the Society Islands hotspot

A similar conclusion may be drawn from consideration of a recorded sequence of Plio-Pleistocene reversals obtained from lavas extruded from the Society Islands hotspot in French Polynesia^{18,19,30}. Records from Tahitian lavas of the R-N Matuyama-Brunhes, the N-R Jaramillo-Matuyama (0.90 Myr) and the closely spaced pair of reversals bounding the Cobb Mountain normal subchron (1.1 Myr)¹⁸ each contain a particular cluster of transitional palaeodirections which are remarkably similar to one another. In contrast to the findings from the site of the Hawaiian hotspot, the corresponding clusters of VGPs are found, not near South America, but near southwest Australia (Fig. 2), nearly 180° away in longitude.

Results from a Pliocene transitional event (3.0 Myr) obtained from the Society Islands hotspot lavas exposed on Huahine Island^{19,30} attest to the significance of the observed orientation of the clustered Pleistocene palaeodirections. Although this very detailed record contains many intermediate directions whose VGPs appear over much of the globe, Roperch and Duncan¹⁹ emphasize the correlational importance of a particular palaeofield orientation found at several sites, sometimes in thick sequences of lava. In fact, this particular direction corresponds to a VGP near northwest Australia (Fig. 2), near to those associated with the Pleistocene recordings.

Comparison with records from sediments

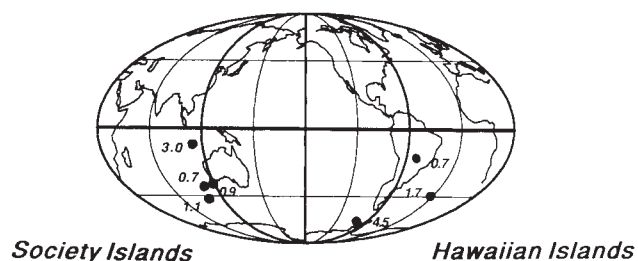
All three Hawaiian VGPs shown in Fig. 2 reside within a 60° band of longitude centred near 310° E. This range overlaps the longitude band claimed to contain an inordinate number of transitional VGP paths⁸⁻¹⁰ obtained primarily from sediments. Also implicated by these sediment studies^{9,10} is a second 'preferred' longitudinal band for VGP paths antipodal to the first, centred along a meridian near western Australia. The VGP clusters recorded at the Society Islands hotspot lie well within this band.

Tric *et al.*⁸ argue that a clear prevalence of available palaeomagnetic reversal records derived from sediments³¹⁻³³ is consistent with dipolar transition fields. Assuming that these data accurately reflect the direction of the reversing field, they could be explained by a simple model in which an equatorial dipole dominates when the axial dipole weakens through zero and subsequently grows with opposite sign. Such a sequence of events would, indeed, produce longitudinal confinement of the true path of the geomagnetic pole. A reversed equatorial dipole would produce the antipodal path. If so, then contemporaneous records from lavas would also be expected to show similar longitudinal confinement of the VGP. Such confinement, however, is seldom observed.

I argue that it is the positive correlation between the proposed locations of the sediment-derived VGP bands⁸⁻¹⁰ and particular recurring lava-derived VGP clusters which supplies a subtle insight into the reversal process. Specifically, the statistically based observation from these sediment records of longitudinal confinement of the VGP may reflect, not the actual movement of the geomagnetic pole during a continuous dipolar transition process, but rather the effect of directional bias imposed on the magnetization of certain sediments. In contrast to the snapshot-like recording of the ambient palaeofield by a thin lava, remanence acquisition in sediments involves some time-averaging. Possible effects of such smoothing on record fidelity during polarity transitions have been explored previously³⁴⁻³⁶.

The observation of similar VGP cluster locations in lava-recorded reversals (Fig. 2) suggests that particular field configurations may dominate much of the process. Indeed, these

FIG. 2 VGPs representing palaeodirectional cluster means recorded at (a) the Hawaii hotspot, 19°N, 205°E (refs 16, 17 and this work) and (b) the Society Islands hotspot^{18,19} (18°S 209°E), for the Plio-Pleistocene transitions indicated by approximate age in Myr. The plot is centred on longitude 207°E.



recurring palaeodirections sometimes constitute all or most of the recorded transition (see, for example, refs 16, 18, 25, 26 and Fig. 1). Such a temporal feature during remanence lock-in of sediments would be expected to bias the resultant magnetization to a degree dependent on a complex function of several factors including the physical properties of the bulk sediment, its magnetic mineralogy and alteration history, the deposition rate, the duration of the intermediate field state and changes in field intensity with time. The record of the N-R Gauss-Matuyama reversal obtained from lake sediments in California³⁷ (Fig. 3) may be an intermediate example of this effect. As can be seen, the path of the VGP is longitudinally confined near one of the preferred bands. This path, however, contains a clustering of sequential VGPs found just off the easternmost tip of Brazil.

Dynamo families and synchronicity

Analysis of global field behaviour during reversals is limited by our often complete lack of knowledge about the synchronicity among contemporaneous recordings from different sites. For this reason, entire transition paths are often used to test against simple field geometries defined by one or more spherical harmonics^{9,38-41}. Given that palaeomagnetically recognizable long-lived transitional states may be involved in the reversal process, I take a different approach, following the suggestion made in ref. 14, by addressing the question of equatorial symmetry of transitional fields. More specifically, I propose a test for the presence of each of the two dynamo families^{12,13}: the family associated with fields having equatorial antisymmetry (EA) and the family associated with equatorial symmetry (ES). This theoretically based concept¹² forms the basis of Merrill and McFadden's¹⁵ model in which polarity transitions are initiated when the ES family couples in a critical fashion with the EA family. In principle, consideration of dynamo family content greatly simplifies the analysis of transitional geometries, as any field configuration, however complex, can be reduced to two 'components'. In terms of spherical harmonics, each of the two families is composed of a dipole term and an infinite number of non-dipole terms (see, for example, ref. 13).

The degree of similarity between the clusters of palaeodirections associated with the Hawaii hotspot in the Northern Hemisphere is more than matched by that recorded at the Society Islands hotspot in the Southern Hemisphere. The proper interpretation of these data, however, depends largely on whether the recorded directional clusters are synchronous. The only available data that can actually be synchronous are the clusters associated with the Matuyama-Brunhes reversal, as this transition is the only one available at present from both sites. Nonetheless, the similarity of the clustered directions from records obtained at each of the sites, most notably the three Pleistocene datasets from Tahiti which span 0.4 Myr, suggests that any such finding for one reversal may be applicable to the other reversals as well.

Testing for dynamo family dominance

Given that the dynamo families have fundamentally different geometry, it turns out that under particular conditions one may analyse a given transitional field for possible domination by

either family. The following tests (see Fig. 4) can be performed when synchronous data are available from only two sites, provided that the sites effectively reside along the same longitude and are symmetrically situated about the Equator, one in the Northern Hemisphere and the other in the Southern Hemisphere:

- If a transitional field is entirely made up of terms belonging to the EA family (also called the dipole¹² or primary¹⁴ family), the vector field directions at such a set of sites will themselves be antisymmetric about the Equator. That is, regardless of harmonic content, each site pair will simultaneously experience values of inclination that differ only in sign, as well as values of declination that mirror one another about the north-south line. Furthermore, for the EA family case, contemporaneous VGPs from such a set of sites will reside at the same latitude and differ in longitude by 180°.
- If a transitional field is entirely made up of terms belonging to the ES family (also called the quadrupole¹² or secondary¹⁴ family), the vector field directions at such a set of sites will themselves be symmetric about the Equator. That is, regardless of harmonic content, each site pair will simultaneously experience identical values of inclination (same sign) and declinations that mirror one another about the east-west line. Furthermore, for the ES family case, contemporaneous VGPs from such a set of sites will reside at a common longitude and as mirror images about the Equator.

With these considerations in mind, we now explore the lava-derived data through two alternative models.

Model 1 (Synchronous clusters). Here I assume that the recorded site-dependent cluster localities correspond to a single field state. Given that the two recording sites (the Hawaii and Society Islands hotspots) are only ~37° apart, and that the clustered palaeodirections (and VGPs) are vastly dissimilar (Fig. 2), any suggested long-lived field state at the time must be of higher order than that of a dipole (see Fig. 4a). In Fig. 5, I plot the VGP associated with the Matuyama-Brunhes directional cluster

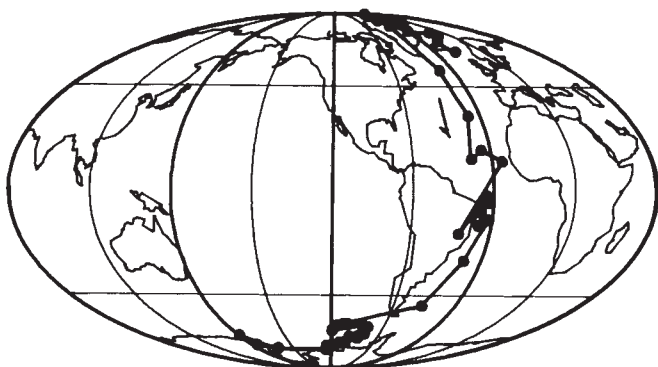


FIG. 3 The path of the VGP associated with the Gauss-Matuyama (N-R) transition as recorded in dry sediments from Searles Lake, California³⁷. The path, strongly confined in longitude, contains a cluster of sequential virtual poles just off the easternmost tip of South America. Quasi-stationarity of the transition field is implied by this cluster, and the recorded confinement of the path may reflect this temporal bias during remanence acquisition.

means recorded both on Tahiti¹⁸ and on Maui¹⁷, along with the VGP predicted for Tahiti, assuming the ambient field to consist solely of EA family members. As can be seen, the observed and predicted VGPs for Tahiti are similar; this is compatible with the possibility that ES family terms decrease with the axial dipole field during reversals. For this case, nondipole terms belonging only to the EA family (such as g_2^1 , g_3^0) would remain to define the stationary field state.

Model II (Nonsynchronous clusters). Here I assume that the recorded site-dependent cluster localities correspond to two transitional field states separated in time. In particular, if one assumes that both transitional field configurations are entirely made up of terms belonging to the ES family, one finds that either directional cluster (or VGP locality) is nearly the reverse of that predicted from knowledge of the other site (see Fig. 4b, c). That is, regardless of the ES family harmonic content (the relative strength of terms such as g_1^1 , g_2^0), when, say, Maui experiences a magnetic field direction corresponding to a VGP near South America, Tahiti would simultaneously experience a direction whose VGP lies essentially along the same meridian at the same angular distance from the Equator in the Northern Hemisphere (this would be near North America). If this ES family field then reverses, one obtains for the site of Tahiti a palaeodirection whose VGP lies within Australia, similar to that actually found in the Tahitian data. Plotted in Fig. 5 is the predicted Matuyama-Brunhes VGP for Tahiti given the recorded palaeodirectional cluster mean on Maui and assuming the global field to consist solely of ES family terms. Also plotted is the antipodal VGP corresponding to the reverse of the predicted direction. As can be seen, this reverse VGP predicted for Tahiti is close to that observed.

Model I calls for the behaviour of the axial dipole term, a member of the EA family, to be somehow linked with the ES family, or at least to act independently from the non-dipole terms of the EA family field. As there is no theoretical justification for such a connection, my preference would be to rule out this model as unrealistic. In contrast, model II suggests that

the clustered transition data from the Hawaii and Society Islands hotspots can be explained by a core process in which the EA family field (and, in particular, the axial dipole) vanishes independently of the ES family field. That is, a complete polarity transition may occur by any process represented by

$$(A, S) \rightarrow (-A, -S)$$

where $\pm A$ and $\pm S$ represent the 'full-strength' states of the EA and ES families, respectively, in which $(0, \pm S)$ intermediate field states are realized.

Model II supposes that lavas extruded at one site, say the Hawaii hotspot, recorded only the $(0, +S)$ field state, whereas those extruded at the other site, say the Society Islands hotspot, recorded only the $(0, -S)$ field state—at least for the polarity transitions considered here, for which one or both palaeomagnetic recordings are available. As a total of six Plio-Pleistocene reversals are represented by these data, such a situation is admittedly improbable. The fact that both hypothesized long-lived field states were not clearly recorded in the same lava sequences may attest to the sporadic and 'clustered' nature of volcanic eruptions. Perhaps one of the two configurations temporally dominates a given event. Yet regardless of whether the principal prediction of model II is correct, namely, that the two dynamo families reverse independently, the inescapable alternative to any two-field-state interpretation of the recorded cluster localities shown in Fig. 2 is that polarity transitions involve single long-lived field geometries having, at most, an insignificant dipole component.

This work has so far incorporated contemporaneous data from only one equatorially mirrored pair of recording sites. Given that this is an early stage in dynamo-family analysis of transitional fields, any particular solution, including model II (the present model of choice), must be considered speculative. Nevertheless, I feel that the inherent simplicity of the interpretation warrants further exploration: for example, a more realistic view of the lava-derived data may be that ES-rich family fields, rather than ES-only family fields, constitute the transitional field

FIG. 4 Schematic diagrams showing the fundamental geometric properties of the two dynamo families on a whole-Earth projection. A pair of low-latitude mirror-imaged sites are used to illustrate directions and corresponding positions of the VGP. (Note: the situation depicted was chosen to be relevant to the observations under discussion.) UP and DN indicate upward (negative) and downward (positive) inclinations, respectively, the magnitude being common to both sites. *a*, Field directions illustrating an EA (equatorially antisymmetric) family configuration. For any such site pair the associated VGPs have the same latitude and differ in longitude by 180° . *b*, Field directions at the same pair of sites chosen to illustrate an ES (equatorially symmetric) family configuration. For any such site pair the associated VGPs are mirror-images about the equator, that is, have the same latitude apart from sign and have the same longitude. (Note: (1) this particular ES family field was constructed by simply reversing the field vectors shown in *a* throughout the Southern Hemisphere. See text for further discussion. (2) As there are an infinite number of possible EA (and ES) family configurations corresponding to the directions shown in *a* and *b*, respectively, meaningful information about possible equatorial symmetry (or antisymmetry) of a given field can only be gained by comparing sites mirror-imaged about the Equator, unless a particular harmonic content is assumed.) *c*, Field directions and VGPs at the same pair of sites associated with the reverse of the ES field in *b*, for comparison. Given the field direction (and associated VGP) shown for the Northern Hemisphere site in *a* and *b*, the field direction (and associated VGP) shown for the mirror-imaged Southern Hemisphere site for the EA family case (*a*) cannot be distinguished from the reversed ES family case (*b*) unless the synchronicity of the two directions is known. That is, the single field state EA and the two field states $\pm$ ES illustrate the synchronous and non-synchronous alternatives, respectively, discussed in the text.

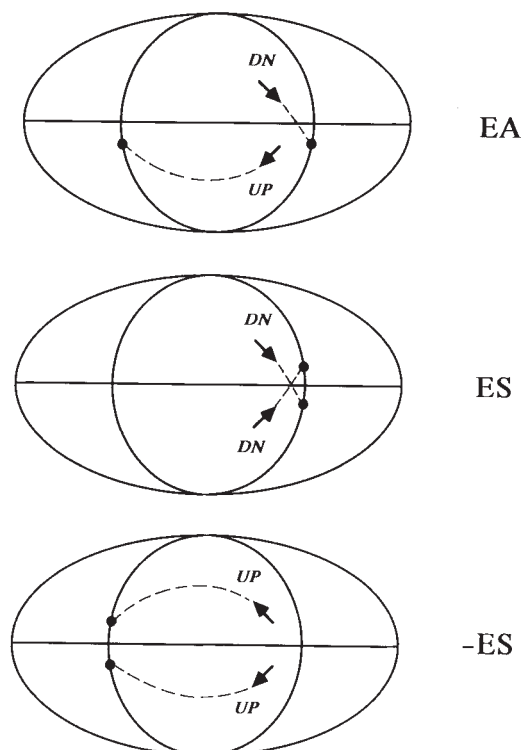
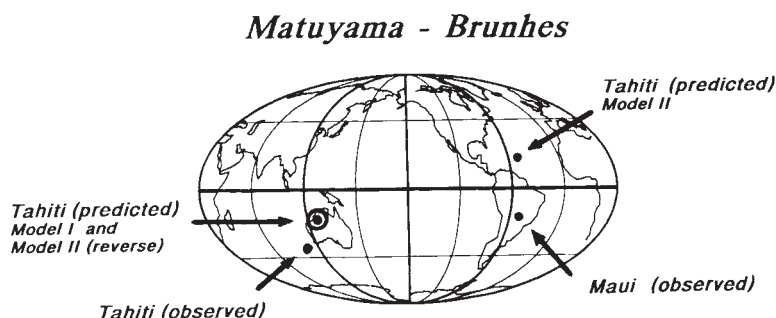


FIG. 5 Matuyama-Brunhes VGP's associated with the cluster mean palaeodirections recorded on Tahiti at the Society Islands hotspot ($D=229^\circ$, $I=2^\circ$)¹⁸ and on Maui at the Hawaii hotspot ($D=105^\circ$, $I=-30^\circ$)¹⁷. Also plotted are VGP's predicted for Tahiti from the Maui data assuming model I, a purely EA family field ($D=255^\circ$, $I=30^\circ$), and assuming model II, a purely ES family field ($D=75^\circ$, $I=-30^\circ$) and its reverse (−ES state) direction ($D=255^\circ$, $I=30^\circ$). (Note that the reversed direction predicted by model II and, hence, its VGP (open symbol) is identical to the nonreversed direction and VGP predicted by model I; see also Fig. 4.)



configurations. In any case, model II is compatible with the contention that long-lived transitional field states may be significantly dipolar. Specifically, it is reasonable to suspect that the equatorial dipole term—the ES family member having lowest harmonic degree and, therefore, the one that weakens least with distance from the core—would dominate an ES-rich field at the Earth's surface. Although the significance of the EA family cannot be determined from these first data, the inclination of the transitional geocentric dipole, a parameter which may be palaeomagnetically recognizable given sufficient data, would provide such an estimate. It should also be noted that the axial quadrupole term (g_2^0) is a member of the ES family. The strength of this term, relative to that of the equatorial dipole, may largely determine the degree to which an ES-rich transition field is axisymmetric. This dichotomy may also explain why some transitional field analyses have involved a rough axisymmetry^{39–42}.

Two dynamo sources?

Theoretically, confirmation of the existence of ES-rich transitional field states need not be evidence for spatial separation of

the principal sources of the two dynamo families¹⁵. Nevertheless, when extrapolated down to the core–mantle boundary⁴³, determinations of power as a function of harmonic degree for the geomagnetic field show the axial dipole term (g_1^0), the lowest degree member of the EA family, to be considerably stronger than the higher-degree harmonics. This finding is compatible with the contention that the EA family field dominates in the deep core. Furthermore, that the ES family field seems to be largely associated with westward drift¹⁵, and the EA family field does not⁴⁴, is consistent with the ES family field source having preferential residence in the shallow core⁴⁵. If so, then polarity reversal may involve the occasional 'destruction' of the deep-core-generated (EA-rich) field, perhaps, through interfamily interactions¹⁵, leaving primarily the shallow-core-generated (ES-rich) field to dictate the intermediate field geometry and its degree of stationarity. The observation of particular long-lived field configurations which recur during transitional events over timescales of, perhaps, several million years, suggests that this geometric control is associated with similarly long-standing lateral variations of the core–mantle system, possibly related to temperature⁴⁶. □

Received 29 April; accepted 7 November 1991.

- Bogue, S. W. & Hoffman, K. A. *Rev. Geophys.* **25**, 910–916 (1987).
- Hoffman, K. A. *Scient. Am.* **256**, 76–83 (1988).
- Laj, C., in *The Encyclopedia of Solid Earth Geophysics* (ed. James, D. E.) 535–547 (Van Nostrand Reinhold, New York, 1989).
- Hoffman, K. A. in *The Encyclopedia of Solid Earth Geophysics* (ed. James, D. E.) 547–555 (Van Nostrand Reinhold, New York, 1989).
- Clement, B. M. & Constable, C. G. *Rev. Geophys. Suppl. U.S. Nation Rep. to Int. Un. Geodesy Geophys. 1987–1990*, 343–348 (1991).
- Watkins, N. D. *Geophys. J. R. astr. Soc.* **17**, 121–149 (1969).
- Creer, K. M. & Isipir, Y. *Phys. Earth planet. Int.* **2**, 115–127 (1970).
- Tric, E. et al. *Phys. Earth planet. Int.* **63**, 319–336 (1991).
- Clement, B. *Earth planet. Sci. Lett.* **104**, 48–58 (1991).
- Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).
- Bogue, S. W. *Nature* **351**, 445–446 (1991).
- Roberts, P. H. & Stix, M. *Astr. Astrophys.* **18**, 453–466 (1972).
- McFadden, P. L., Merrill, R. T. & McElhinny, M. W. *J. geophys. Res.* **93**, 11583–11588 (1988).
- Merrill, R. T. & McFadden, P. L. *Science* **248**, 345–350 (1990).
- Merrill, R. T. & McFadden, P. L. *J. geophys. Res.* **93**, 11589–11598 (1988).
- Bogue, S. W. & Coe, R. S. *Nature* **295**, 399–401 (1982).
- Coe, R. S., Hsu, V. & Theyer, F. (abstr.) *Eos* **66**, 872 (1985).
- Chauvin, A., Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2727–2752 (1990).
- Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2713–2726 (1990).
- Kirschvink, J. L. *Geophys. J. R. astr. Soc.* **62**, 699–718 (1980).
- Hoffman, K. A. *J. geophys. Res.* **89**, 6285–6292 (1984).
- Tang, C. & Coe, R. S. *SEDY Symp. Abstr. Santa Fe, New Mexico* (1990).
- Bogue, S. W. & Coe, R. S. *J. geophys. Res.* **89**, 10341–10354 (1984).
- Prevot, M., Mankinen, E. A., Gromme, C. S. & Coe, R. S. *Nature* **316**, 230–234 (1985).

- Shaw, J. *Geophys. J. R. astr. Soc.* **40**, 345–350 (1975).
- Hoffman, K. A. *Nature* **320**, 228–232 (1986).
- Laj, C., Guitton, S. & Kissel, C. *Nature* **330**, 145–148 (1987).
- Hoffman, K. A. *Nature* **309**, 90–91 (1984).
- Valet, J.-P. & Laj, C. *Nature* **311**, 552–555 (1984).
- Roperch, P. & Chauvin, A. *Geophys. Res. Lett.* **14**, 151–154 (1987).
- Valet, J.-P., Laj, C. & Langeris, C. G. *J. geophys. Res.* **93**, 1131–1151 (1988).
- Herrero-Bervera, E., Theyer, F. & Hellsley, C. E. *Phys. Earth planet. Int.* **49**, 325–342 (1987).
- Clement, B. M. & Kent, D. V. in *Init. Rep. Deep Sea Drilling Project 94*, 831–852 (1986).
- Hoffman, K. A. & Slade, S. *Geophys. Res. Lett.* **13**, 483–486 (1986).
- Constable, C. G. *J. geophys. Res.* **95**, 4587–4596 (1990).
- Rochette, P. *Earth planet. Sci. Lett.* **98**, 33–39 (1990).
- Liddicoat, J. C. *Phil. Trans. R. Soc. Lond. A306* 121–128 (1982).
- Hillhouse, J. & Cox, A. *Earth planet. Sci. Lett.* **29**, 51–64 (1976).
- Hoffman, K. A. *Science* **196**, 1329–1332 (1977).
- Hoffman, K. A. & Fuller, M. *Nature* **273**, 715–718 (1978).
- Williams, I. & Fuller, M. *J. geophys. Res.* **86**, 11657–11665 (1982).
- Weeks, R. J., Fuller, M. & Williams, I. *J. geophys. Res.* **93**, 11613–11620 (1988).
- Lowe, F. J. *Geophys. J. R. astr. Soc.* **36**, 717 (1974).
- Gubbins, D. & Bloxham, J. *Nature* **325**, 509–511 (1987).
- Bullard, E. C., Freedman, C., Gellman, H. & Nixon, J. *Phil. Trans. R. Soc. Lond. A243*, 67–92 (1950).
- Gubbins, D. *Nature* **326**, 167–169 (1987).

ACKNOWLEDGEMENTS. I thank R. Holcomb for sharing his knowledge about Molokai and the recorded reversal boundary, D. Morse for assistance with field and laboratory work, R. Coe for supplying unpublished data from Maui, R. Merrill and P. McFadden for discussions and M. Prevot for written comments. D. Gubbins suggested renaming the two dynamo families on the basis of equatorial symmetry. This project was supported by the NSF.

Structure of simian virus 40 at 3.8-Å resolution

R. C. Liddington^{*†‡}, Y. Yan^{*†}, J. Moulaï^{†‡}, R. Sahli^{§‡}, T. L. Benjamin[§]
& S. C. Harrison^{*†}

^{*} Howard Hughes Medical Institute and [†] Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA
[§] Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

The crystallographically determined structure of simian virus 40 shows that the 72 pentamers of viral protein VP1, which form the outer shell, have identical conformations except for the C-terminal arms of their subunits. Five arms emerge from each pentamer and insert into neighbouring pentamers. This tying together of standard building blocks allows for the required variability in packing geometry without sacrificing specificity.

THE polyomaviruses, of which simian virus 40 (SV40) and murine polyoma are examples, are the simplest of the viruses with double-stranded DNA genomes¹. They can induce tumours in laboratory animals and transform cells in culture¹. The virus particles (about 500 Å in diameter) deliver a minichromosome from the nucleus of one cell to the nucleus of another. They are larger than the positive-strand RNA viruses (about 300 Å in diameter) that have been studied crystallographically^{2,3}, and they pose different problems in assembly and intracellular localization. They are icosahedrally symmetrical particles⁴, containing 360 copies of a major structural protein, VP1 (of relative molecular mass about 40,000; $M_r \sim 40K$), arranged as 72 pentameric building blocks on the viral surface⁵. They also contain smaller quantities of two internal proteins, VP2 and VP3 (ref. 1).

The organization of polyomavirus capsids presents a structural puzzle^{5,6}. We can distinguish two classes of VP1 pentamers, according to their location in the shell (Fig. 1). Twelve 'white' pentamers lie on the twelve 5-fold rotation axes of the icosahedron, each surrounded by five other pentamers; the remaining 60 'coloured' pentamers do not lie on symmetry axes, and they are surrounded by six other pentamers (one white and five coloured). The puzzle is how do the coloured pentamers fit into hexavalent holes? What are the elements of flexibility and alternative bonding that allow for the mismatch of symmetries?

This unexpected organization was first discovered in the low-resolution structure of murine polyomavirus^{5,6}. They observed that the particular relative orientations of the 5- and 6-coordinated pentamers lead to just three kinds of interpentamer contacts: an approximate 3-fold axis relating subunits labelled α , α' and α'' ; an approximate 2-fold axis relating those labelled β and β' ; and a strict icosahedral 2-fold axis relating those labelled γ . It was proposed that correct assembly required 'switching' among the alternative modes of interpentamer bonding, but the resolution was insufficient to define its chemical nature.

We report here the structure of SV40 to 3.8-Å resolution, determined using X-ray crystallography with synchrotron radiation. The trace of the VP1 polypeptide chain shows a striking, and potentially quite general, way of achieving specific interactions between subunits with alternative relative orientations. The core of a VP1 pentamer has an essentially invariant 5-fold

structure, defining a uniform building block for the particle. Long C-terminal arms tie the pentamers together; they extend from one pentamer and fit into binding sites on adjacent pentamers. Variability in the geometry of contacts is accommodated almost entirely by the different directions that an arm can take as it emerges from the core of its subunit and heads toward a neighbour. Its interactions when inserted in the binding site on that neighbour are again essentially invariant. This is quite different from what is found in the RNA virus structures. Our structure also accounts for a number of other aspects of virion stability and biochemistry, including specific incorporation of the internal proteins.

Structure determination

The X-ray data collection is summarized in Table 1. The highest resolution data were obtained using short-wavelength (0.88-Å)

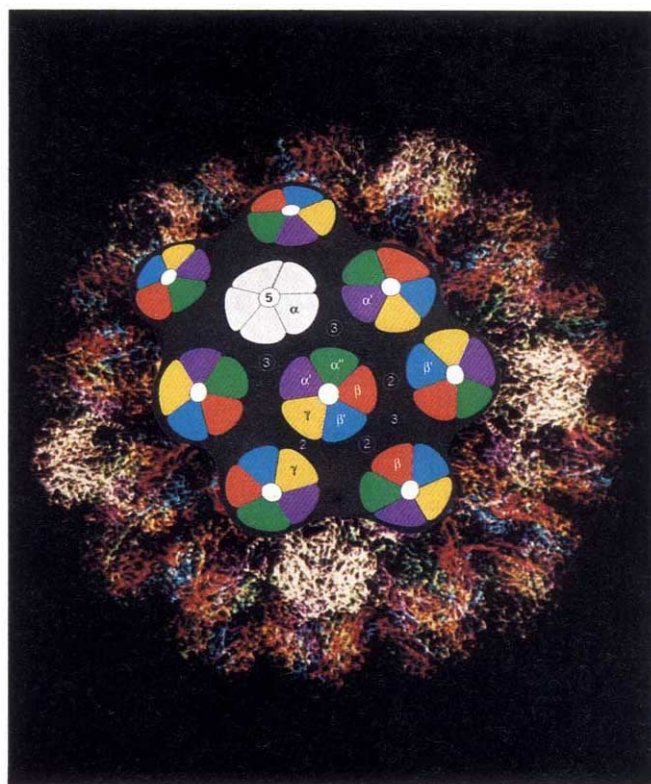
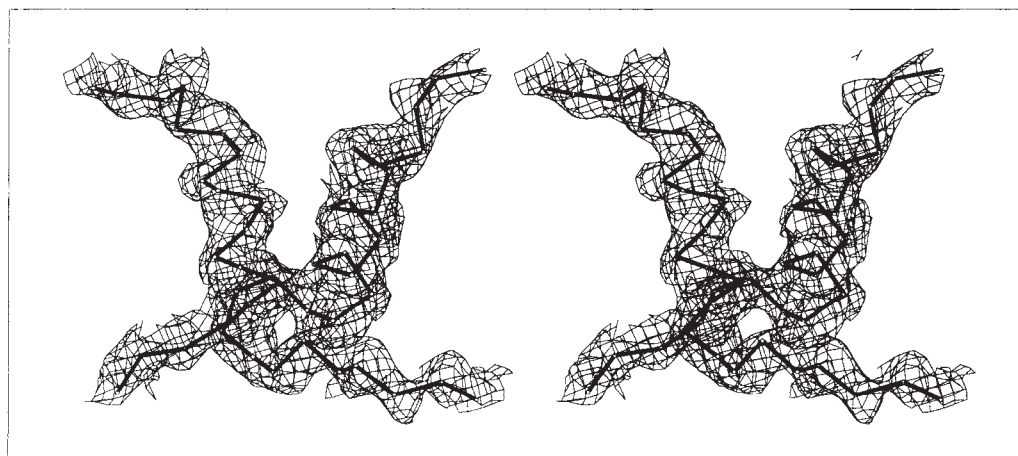


FIG. 1 Overall view of the SV40 virion. The shell is composed of 360 copies of VP1, organized into 72 pentamers. The pentamers are situated at the points of a $T = 7d$ icosahedral surface lattice. There are twelve 5-coordinated pentamers (white) and 60 6-coordinated pentamers (coloured). The three kinds of interpentamer clustering are indicated in the schematic part of the diagram. The white (α), purple (α') and green (α'') subunits form a 3-fold interaction (③); the red (β) and blue (β') subunits form one sort of 2-fold interaction (②); the yellow (γ) subunits form another sort of 2-fold interaction (2). Icosahedral symmetry axes are indicated by numbers: 5, 3, 2.

[‡] Present addresses: Dana Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA (R.C.L.); Brandeis University, 415 South Street, Waltham, Massachusetts 02154, USA (J.M.); Institute of Microbiology CHUV, Lausanne, Switzerland (R.S.).

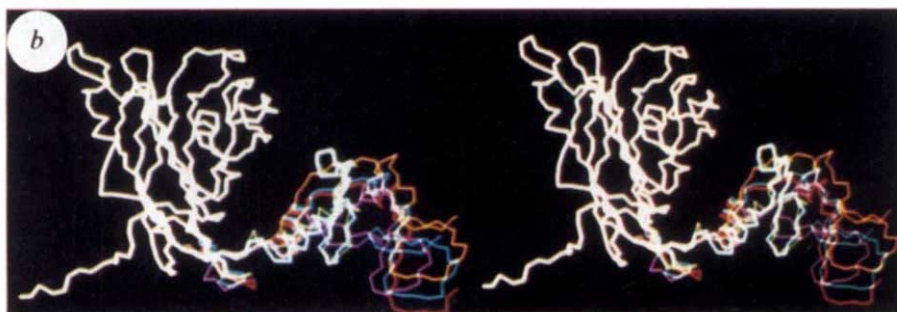
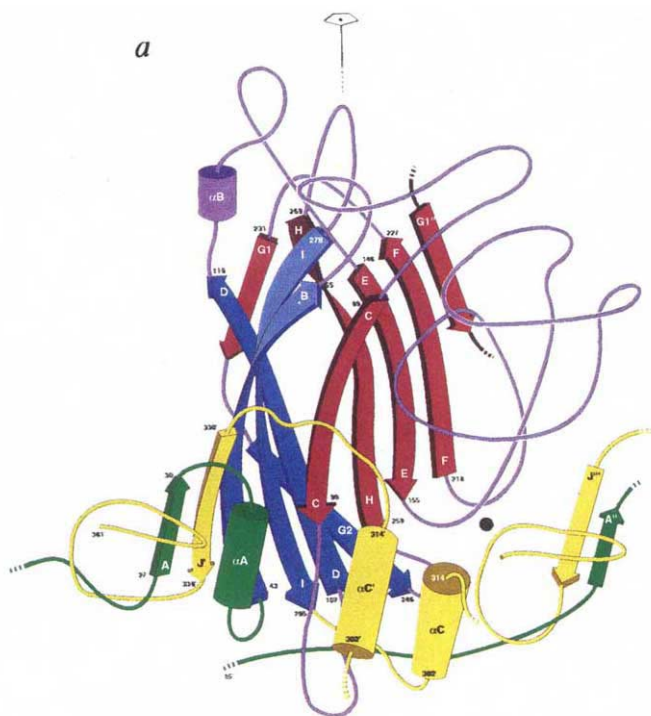
FIG. 2 Portion of the final SV40 electron density map at 3.8-Å resolution, after icosahedral phase refinement as described in the text. The region shown includes two C helices in a $\beta:\beta'$ (red:blue) interaction, as well as parts of the chains leading into and out of these helices. The map demonstrates how the C-terminal arms emerge from their respective subunits, pass by each other, form C helices and extend into the opposite subunit.



radiation. For phase determination, we used a single heavy-atom derivative (methyl mercury nitrate), combined with noncrystallographic symmetry phase refinement. We used a Patterson sum search procedure^{7,8} to locate the mercury atoms. Although the search map was noisy, a pentagonal arrangement of peaks could be discerned around the presumed axis of the hexavalent pentamer. These positions were refined with the program HEAVY, using origin-removed difference Patterson coefficients⁹. A difference Fourier map revealed additional peaks related by this local 5-fold axis and showed related peaks on the pentavalent pentamer. A total of 20 sites per pentamer (120 sites per crystallographic asymmetric unit) were found by successive difference Fourier maps, from which single isomorphous

replacement (SIR) phases were calculated. These phases were used to construct an SIR map of the virus at 6.5 Å; icosahedral 5-fold averaging produced a very clear map from which we drew a simple envelope, consisting of a spherical shell between radii of 160 and 220 Å, and projecting towers, 50 Å high and 90 Å across, centred on each pentamer. Icosahedral phase refinement, using the 5-fold, noncrystallographic redundancy¹⁰⁻¹², produced

FIG. 3 *a*, Richardson diagram of the SV40 VP1 subunit. Capital letters, elements of secondary structure; numbers, first and last residues of each strand. Two C helices are included, one belongs covalently to the subunit in the diagram, the other belongs to the invading arm (primed letters and numbers). The rest of the invading arm is also shown, but not the rest of the arm of the central subunit. An N-terminal segment from a neighbouring subunit in the same pentamer (the anticlockwise neighbour) appears (doubly primed numbers), as well as a C-terminal segment from the arm that invades the neighbour (triply primed numbers). The solid dot shows the proposed Ca^{2+} site. The colours are chosen as follows: blue, BIDG2 sheet; red, CHEFG1 sheet; violet, loops connecting strands of β sheets; green, N-terminal arm and clamp; yellow, C-terminal arm. Note that the central subunit in this drawing interacts with the N-terminal arm from the anticlockwise neighbour in the same pentamer and with C-terminal arms from two other pentamers. *b*, Stereo drawing showing the six conformations of VP1. Residues 15–295 are essentially invariant, except for the CD loop (99–106). The C-terminal arms all have essentially the same structure, but they differ in the direction they take as they enter and leave the C helix and in the orientation of the C helix itself.

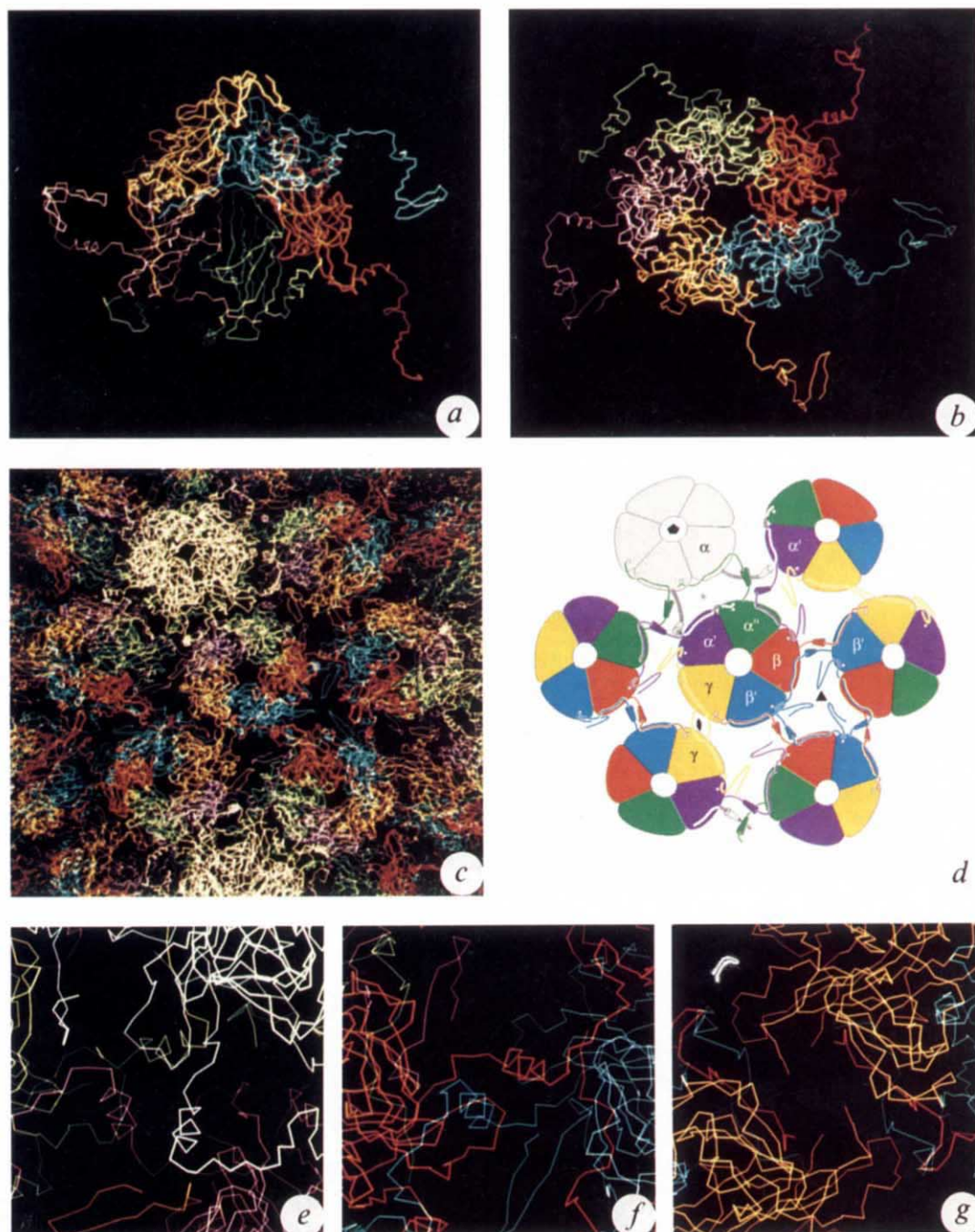


a map in which helices and sheets were readily apparent. Phase extension¹³ to 4.5 Å was done in two steps (6.5–5 Å and 5–4.5 Å), starting from a phase set comprising the refined 6.5-Å set and the limited high-resolution SIR phases. This higher-resolution map revealed sensible new features such as resolved strands in the β sheets. We next tried phase extension in steps of two reciprocal-lattice units, using an improved envelope and 5-fold averaging; to the limit of the native data (3.8 Å), but despite reasonable final statistics this procedure did not noticeably improve the map. Careful comparison of pentamers revealed that the cores of the pentavalent and hexavalent pentamers had identical structures, and that the only variable parts were around the bases of the pentamers in the regions of pentamer-pentamer contact. We therefore repeated the phase extension from 4.5 to 3.8 Å, this time averaging the density for both pentamer cores (30-fold averaging for VP1), and 5-fold averaging the variable parts. The matrix to relate the hexavalent

pentamer to the pentavalent pentamer was derived from the heavy atom positions; the root-mean-square (r.m.s.) deviation for a least-squares fit of the mercury atoms on the hexavalent and pentavalent pentamers was only 0.4 Å. This map (Fig. 2) showed much improvement over the previous map, and a preliminary tracing of the chain began. The mercury sites labelled the positions of cysteine residues (four sites per VP1); sequence comparison between SV40 and polyoma VP1 suggested the positions of loops; the resolution of the map was sufficient to distinguish between large and small side chains. Still some ambiguities remained, particularly in the C-terminal arms, which were resolved by collecting data from crystals soaked in K_2PtCl_4 , which labelled several methionines. At a later stage, the Gd^{3+} derivative confirmed the location of three acidic groups.

Except for the N-terminal 15 residues, the complete polypeptide chain has been built. For the β -sandwich framework, we are confident of the main-chain locations and the side-chain

FIG. 5 Views of the VP1 pentamer and of its interactions in the virion. *a*, A 6-coordinated pentamer, with its full complement of arms represented as they are configured in the virion, tipped so that it is viewed from the inward-facing surface. Note the conical hollow along its axis. *b*, As in *a*, viewed from the outside of the virion. *c*, Closeup of the virion surface, viewed along the axis of a 6-coordinated pentamer. *d*, Schematic drawing, showing the way C-terminal arms tie the pentamers together. The view corresponds to *c*. The body of a pentamer is represented by a five-petalled flower; the arms are represented by cylinders and lines. Arms emanate from a subunit, form C helices (small cylinders) and C inserts (lines running through the invaded subunit), and in three cases also form ordered C loops (hook-like loops filling gaps between pentamers). The colour code is the same as in Fig. 1. Icosahedral symmetry axes are shown by the conventional symbols ($\blacklozenge$, $\blacktriangle$, $\blacklozenge$, for 5-, 3- and 2-fold, respectively). The asterisk shows a local 2-fold symmetry axis relating the 5- and 6-coordinated pentamers. The closed circles mark the locations of proposed Ca^{2+} sites: note that they seem to contribute to the anchoring of an arm in its target pentamer. *e*–*g*, Details of the three types of interpentamer cluster, magnified from *c*.



directions. For the connecting elements, especially parts of the large EF loop, there is uncertainty in many side-chain directions, and the sequence could be out of register in some places by one or perhaps two residues.

Subunit structure

The tertiary structure of the VP1 divides the polypeptide chain into three modules: an N-terminal arm; an antiparallel β -sandwich with jelly-roll topology; and a long C-terminal extension (Figs 3 and 4). The first 15 residues of the N-terminal arm are not visible in the electron-density map, but they probably extend inward to interact with the minichromosome. The central β -sandwich can be described with the same strand labelling (B-I) as the RNA virus capsid domains of picornaviruses and plant viruses^{2,3}, but the lengths of the strands and the relative positions of the two sheets impart a quite different overall shape. The strands run roughly parallel to a 5-fold axis. Elaborate loops emanate from the β -sheet framework. The BC, DE and HI loops form closely interacting structures at the outward facing end of the β -sandwich. The long EF loop (64 residues), which originates and inserts at the inward facing end, forms a small jelly-roll structure on the side of the β -sandwich. Two strands augment the BIDG sheet: four residues of the N-terminal arm form strand A, and the C-terminal arm of a subunit in an adjacent pentamer contributes strand J. The A strand and the short A helix form a 'clamp' that locks this J strand in place. The C-terminal arm can be subdivided into three segments: the 'C helix' (residues 302–314), the 'C insert' (residues 315–344), and the 'C loop' (residues 345–361). Residues 330 to 336 of a C insert make a J strand.

The structure of the VP1 β -sandwich is the same in all six unique environments. The only significant variation occurs in the C-terminal arms and the CD loops (Fig. 3b).

Pentamer structure

Five VP1 subunits are intimately associated in a pentamer (Fig. 5a, b). Even the secondary structures interlock. The G strand is

in two parts: the C-terminal part (G2) makes several sheet H-bonds in the BIDG sheet of the subunit from which it emanates; the N-terminal part (G1) contributes a fifth strand to the CHEF sheet of the clockwise neighbour. The D strand leaves the BIDG sheet halfway up; it then makes an extensive DE loop, including a turn of α helix that nestles into the small jelly roll (EF loop) and the BC and HI loops of the clockwise neighbour. The pentamer is roughly cylindrical, about 80 Å in diameter and 70 Å tall, extending from a radius of 180 to 250 Å in the virus particle. It has a hollow conical interior of diameter roughly 50 Å at its base, tapering to about 12 Å at a neck formed by the short FG corners (Fig. 5a). This neck opens up on the top surface into a dimple about 20 Å wide and 12 Å deep. The large BC, DE and HI loops from each subunit interlock to form the top surface of the pentamer. The side of the pentamer is formed by the outer edge of the sandwich, the EF loop, the clamp and the invading arms. Residues that face the conical hollow of the pentamer come largely from the F and G strands. Because of the excursion made by the G strand away from its own subunit, the inner edge of the sandwich is partly exposed, so that a few sidechains from the D and E strands also contribute to the inner surface.

The N-terminal arm extends across the inside of the pentamer beneath the clockwise neighbouring subunit (Fig. 5a). The first ordered residue of the arm, Lys 15, contacts the connection between the A helix and the B strand of the neighbour, around residue 40. The N-terminal sequence contains the VP1 nuclear localization signal¹⁴, and mediates DNA-binding *in vitro*¹⁵. It is completely internal in a virus particle, and it therefore cannot contribute to nuclear transport of an intact virion (J. Clever, M. Yamada and H. Kasamatsu, manuscript in preparation). It is, of course, exposed on a free pentamer.

Pentamer-pentamer contacts in the virion

The most striking aspect of the assembled virion is the disposition of the C-terminal arms (Fig. 5c, d). These arms form the principal interpentamer contacts, by extending away from their

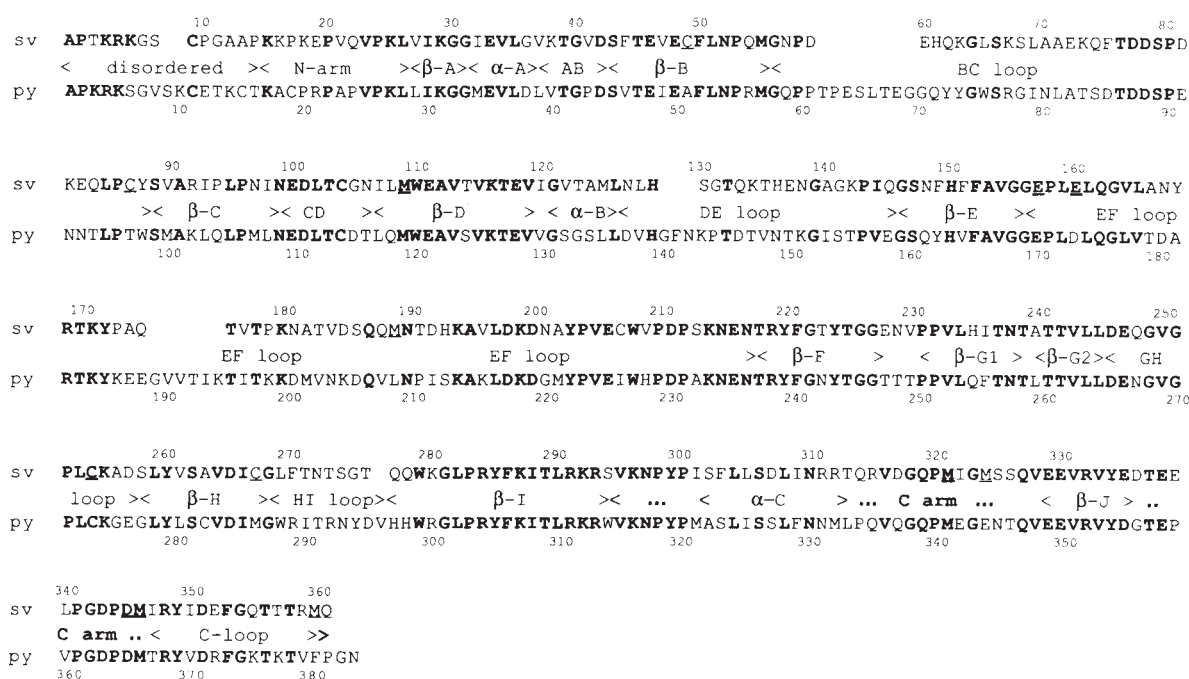


FIG. 4 Sequences of VP1 from SV40 and murine polyoma (see ref. 1), with secondary-structure assignment from the model reported here (single-letter amino-acid code). Underlined residues in the SV40 sequence show methionines

labelled by K_2PtCl_4 (M), cysteines labelled by CH_3HgNO_3 (C), and acidic ligands for Gd^{3+} (D, E).

TABLE 1 X-ray data collection

	Overall	20–11.2	11.2–7.9	7.9–6.5	Resolution ranges (Å)					
					6.5–5.6	5.6–5.0	5.0–4.6	4.6–4.2	4.2–4.0	4.0–3.8
Data completeness	0.90	0.99	0.99	0.99	0.99	0.98	0.98	0.95	0.85	0.50
R_{merge}^*	0.24	0.08	0.11	0.23	0.32	0.31	0.32	0.37	0.42	0.41
Phasing power of methyl-mercury derivative†	1.29	1.87	1.85	1.67	1.21	0.97	0.80			
Final statistics of noncrystallographic symmetry phase refinement:										
Averaging R	0.22	0.07	0.07	0.13	0.19	0.19	0.21	0.29	0.37	0.41
Correlation‡	0.89	0.98	0.99	0.96	0.89	0.88	0.86	0.73	0.64	0.70
Total film packs		170								
Total reflections		1,001,317								
Independent reflections		262,283								

Virus was grown, purified and crystallized with only minor modifications to conditions described previously³⁷. Crystals belong to space group I23, $a = 558$ Å; the virus is centred on lattice points of the unit cell, and the 2-fold and 3-fold crystallographic axes coincide with icosahedral 2- and 3-fold axes. One-twelfth of the virion lies in the crystallographic asymmetric unit (one pentavalent pentamer and six hexavalent pentamers). One-fifth of this is the icosahedral asymmetric unit; one subunit of the pentavalent pentamer and one complete hexavalent pentamer. There is an ambiguity in the direction of the icosahedral 5-fold axis with respect to the crystallographic symmetry axes, which was resolved by looking at the directions of the intensity spikes in the diffraction pattern. Data were collected initially at the Cornell high-energy synchrotron source (beamline A1) using a wavelength of 1.5 Å. A complete native data set to 4.5 Å spacing, and a complete methyl mercuric nitrate derivative data set to 6.5 Å (with a little data to 4.5 Å) were collected in 0.5° steps. The crystals are radiation-sensitive, and only one data-quality photograph was obtained from each region of a crystal. Crystals were roughly aligned optically before X-ray exposure. Data were later collected at the Daresbury synchrotron (station 9.6) using a wavelength of 0.88 Å in steps of 0.2°. The short-wavelength radiation allowed us to collect higher resolution data ($d_{\text{min}} = 3.8$ Å) and to take about four times as many films per crystal. The highest resolution data were obtained using short-wavelength (0.88-Å) radiation. For phase determination, we used a single heavy-atom derivative CH_3Hg^+ , combined with noncrystallographic symmetry phase refinement^{11,12} and phase extension¹³. Films were digitized with an Optronics P-1000 densitometer, and intensities were integrated using the program SCANFILM^{38,39}. Data were scaled and merged using the CCP4 program suite. The program ROTAVATA was modified to use high-percentage partial reflections in scaling.

* $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \bar{I}_h| / \sum_h \sum_i I_{hi}$, where h are unique reflection indices, I_{hi} are intensities of redundant, symmetry-related reflections of index h , and $\bar{I}_h$ is the mean intensity for reflections of index h .

† Phasing power, mean value of heavy-atom structure-factor amplitudes divided by the residual lack of closure error.

‡ Averaging $R = \sum_h ||F_o| - |F_c|| / \sum_h |F_o|$, where F_o and F_c are observed and calculated structure factors.

§ Correlation coefficient $= \sum_h (|F_o| - |F_c|)(\langle F_c \rangle - |F_{cn}|) / [\sum_h (|F_o| - |F_{on}|)^2 - (\langle F_c \rangle - |F_{cn}|)^2]^{1/2}$.

subunit of origin and invading a subunit of an adjacent pentamer. Each pentamer receives five invading arms, one from each of five other pentamers, and donates five arms to surrounding pentamers. The pattern of exchange among different pentamers, described in the next section, determines the surface lattice. The C insert (residues 315–344) interacts in all cases identically with the subunit it invades. Residues 315–320 wrap across the outer edge of the β -sandwich, about halfway up its length; residues 321–329 make a loop above the clamp; residues 330–336 form the J strand, which inserts between strands A and B to complete an extended AJBIDG antiparallel sheet; and residues 337–344 wrap around the N-terminal arm and end at the presumed Ca^{2+} -binding site. The C helix and the C loop are the variable parts of the C-terminal arm. They are involved in dimer and trimer interactions, and are disordered on certain of the subunits. Their structures are described further below.

We note that a clamp involving insertion of a strand into an extended β -sheet also exists in the picornaviruses, linking the pentameric subassemblies^{40,41}.

Three types of pentamer cluster

The particular relative orientations of the hexavalent and pentavalent pentamer lead to just three distinct kinds of pentamer-pentamer interactions: a 3-fold cluster and two different 2-fold clusters (Fig. 5c, d). The clusters are determined by interchange of C-terminal arms.

The 3-fold cluster includes one pentavalent pentamer and two hexavalent pentamers. At the centre of this cluster, three C helices, one from each pentamer, associate in an α -helical bundle (Fig. 5e). The angle between helices is about 60°, and they have a hydrophobic contact. The opposite, rather polar face of each helix packs against the CD and GH loops of a subunit in the clockwise neighbouring pentamer. The C insert that emerges from each helix also invades this subunit, as described above. Thus, the pentamers in a 3-fold cluster exchange C-terminal arms in a cyclical fashion. Beneath the

three helices lie the tips of the CD loops, which come into close contact. Local 2-fold axes relate the pentavalent pentamer to each of the hexavalent pentamers in the 3-fold cluster, and further contacts across these 2-folds involve interactions between the A-strand of one apposed subunit and the A helix of the other.

Both kinds of 2-fold clusters contain two hexavalent pentamers. The first kind creates a local 2-fold interaction between pentamers related by the icosahedral 3-fold axis. The participating pentamers exchange C-terminal arms, and the corresponding C helices associate as a two-chain α -helical bundle with a packing angle of about 60° (Fig. 5f). Thus, this α helix provides a surface that can make either a 2-fold or a 3-fold contact. The opposite face of each helix lies against the CD and GH loops of the opposite pentamer, as in the 3-fold cluster, and the tips of the CD loops extend toward each other beneath the paired helices.

The second kind of 2-fold cluster contains pentamers related by an icosahedral 2-fold axis. The pentamers dock so closely that there is no space for the C helices, and electron density for the chain disappears (Fig. 5g). Density reappears as a β strand, which makes a seventh strand of an AJBIDG sheet on one side and interacts with an A helix on the other. The pentamers probably exchange C-terminal arms, but in the absence of continuous electron density, we must also allow the possibility that the C-terminal arms are in this case not extended, but wrapped back into their own subunits.

Fate of the C-terminal hairpin loops

The last 18 residues of the C-terminal arm form a hairpin loop that is ordered to differing extents in various subunits (Fig. 5b–d). The role of this loop seems to be simply one of plugging up holes in the shell. The tips of three loops meet at the icosahedral 3-fold axis; the tips of two loops meet halfway between the icosahedral 2-fold axis and the 3-fold helix cluster. The loops are disordered where pentamers dock closely, at the

icosahedral 2-fold and at the local 2-fold between pentavalent and hexavalent pentamers.

Metal-binding site

Because of the apparent significance of Ca^{2+} for virion stability^{16,17}, we used difference maps to search for possible binding sites. A map calculated using differences between native data and data from crystals soaked in EGTA showed no significant features. Crystals subsequently soaked in Ca^{2+} -free harvest buffer containing 1 mM $\text{Gd}(\text{NO}_3)_3$ produced a difference map with clear peaks at four out of six chemically similar positions. We believe that the high concentration of ammonium sulphate (2–3 M) used to grow and stabilize these crystals may have caused sequestration of bound Ca^{2+} as insoluble CaSO_4 , but that the greater solubility of $\text{Gd}_2(\text{SO}_4)_3$ has allowed Gd^{3+} to bind to a Ca^{2+} site, as it does in tomato bushy stunt virus (TBSV)⁴². The bound ion forms a bridge between a C-terminal arm of one pentamer and an EF loop of another. Two glutamate residues (157 and 160) from the EF loop and one aspartate residue (345) from the invading C-terminal arm are in the correct location to coordinate the metal. All three acidic residues are conserved in known polyomavirus VP1 sequences. The EF loop providing the glutamates is not in the subunit directly invaded by the arm providing the aspartate, but rather in the clockwise neighbour. The ion bridge thus provides an additional link that fastens an arm into a target pentamer.

Disulphides?

Reducing agents are required to disassemble polyoma virus¹⁸ and SV40¹⁷ *in vitro*. It has therefore been proposed that disulphide bonds stabilize the virions and that reduction of these bonds is a critical step in uncoating. There are no disulphides in the SV40 crystal structure in a subunit or in a pentamer, but it is possible that disulphide formation occurs between CD loops of neighbouring pentamers (Cys 104). These loops approach each other just beneath the 2-fold and 3-fold clusters of C helices. They contain at their point of closest approach one of only two cysteines conserved among polyomaviruses (the other is on the GH loop, and it binds Hg in our derivative). The density is very weak near the CD-loop cysteines, however. In the case of the 3-fold interaction, statistical disorder could explain the poor density, as all three sulphhydryls cannot participate simultaneously. Flexibility of the loops is the only explanation for weak density in the 2-fold interaction, but a disulphide link may still be present.

Internal proteins

The high-resolution electron density map shows a clear density feature in the hollow of each pentamer, just beneath the constriction formed by the FG corners of the five subunits. It effectively plugs the hole along the pentamer axis. Only the N-terminal 15 residues of VP1 are not accounted for in our chain trace. They project toward the inside of the virus, and they cannot readily reach back into the pentamer hollow. We therefore believe that the axial feature, which is present on both the pentacoordinated and hexacoordinated pentamers, represents a segment of VP2 or VP3. There is roughly one chain of an internal protein (VP2 or VP3) per VP1 pentamer in an SV40 virion. It is reasonable to propose that there is a specific association and that the interaction is mediated by some part of the sequence common to VP2 and VP3. Because all five symmetrically related binding modes probably occur with equal frequency, even in the hexacoordinated pentamer, the visible density feature is an averaged image, and identification of the putative VP2/VP3 segment is not possible. Very-low-resolution difference maps between polyoma virions and empty particles reveal density that fills the conical hollow of each pentamer (J. Griffith *et al.*, manuscript in preparation). The authors of that study conclude, as we do, that this density represents an internal protein. The association of an internal protein with each pentamer is con-

sistent with the view that VP2 and VP3 direct assembly of VP1 around a minichromosome^{19,20}.

Comparison with other polyomaviruses

Alignment of the VP1 sequences from murine, hamster and human polyomaviruses with the sequence of SV40 (ref. 1) shows that the jelly-roll framework is highly conserved and that variability is concentrated primarily in the external loops (BC, DE, FG, HI and the top of the long EF insertion). These are likely to be the structures that mediate cell attachment and to be the principal antigenic determinants. The cellular receptor for SV40 is not known. Attachment of murine polyomavirus requires sialic acid on the host cell surface²¹. A point mutation at residue 91 in polyoma VP1, homologous to residue 81 at the base of the BC loop in SV40, changes polyoma's attachment properties^{22–24}. The BC loop in polyoma is eight residues larger than in SV40, and it is likely to form a shallow sialic-acid binding pocket at the outside rim of the VP1 cylinder.

Virus assembly

Almost all contacts between pentamers involve C-terminal arms. We expect these arms to be flexible and unstructured on a free pentamer. To use a mechanical metaphor, we can say that the pentameric building blocks are tied together with ropes rather than cemented together across extended interfaces. The ropes, which issue from one pentamer, are inserted into a slot on another pentamer and secured by a clamp. Specificity is ensured by the sequence of amino-acid residues in the arm, which make precise contacts in the target subunit, but many restrictions are relieved by not needing to match preformed, complementary surfaces. The interactions of C helices seem to impart rigidity to the final structure. Each C helix packs between one or two other C helices and the β -sandwich of the subunit that its C insert invades; it has no close contacts with the body of its own subunit.

The VP1 of polyomavirus has been expressed in *Escherichia coli*²⁵. It is found entirely in pentameric form, as might be expected from the tight association of subunits in a pentamer. The pentamers can be induced to assemble into virus-like shells, as well as into various other structures^{6,26}. Deletion of the C-terminal arm does not affect pentamer formation but prevents assembly into capsids²⁷. Several variant structures have been analysed, including an *in vitro* assembled particle with octahedral symmetry, as well as tubular assemblies of VP1 from infected cells^{26,28,29}. The variant structures all appear to contain 3-fold and 2-fold clusterings similar to those found in the virion, but differently distributed so as to produce a different overall design. Accurate assembly of a 72-pentamer virus must therefore involve formation of the various clusterings in correct relative positions in a shell.

What determines the choice of clustering when VP1 pentamers associate? The prevalence of 3-fold clusters in the observed polymorphic structures suggests that 3-fold interchange of arms is the preferred association, perhaps directed by the 3-fold bundle of C helices. If so, then assembly might ordinarily begin with formation of a complex of five pentamers around one, which contains five such 3-fold interactions. Addition of further pentamers to this structure can involve 3-fold interactions only if the curvature is strained, leading to a 12-pentamer, $T = 1$ shell. Such particles are observed as products of *in vitro* VP1 assembly²⁶, but they are too small to package the viral DNA. Thus, if assembly is initiated on a viral minichromosome, the pentamers surrounding an initial five-around-one can only make 2-fold interactions by pairwise arm exchange with pentamers at the edge of the growing shell (see Fig. 5c). That is, at least some aspects of the correct design can be achieved by a preference for 3-fold arm interchange, with 2-fold exchange a 'second choice' when packaging restricts the curvature. The pentamer 'decides' what to do on the basis of the geometry of the site at which it joins the shell. Additional constraints may be needed,

however, to avoid mistakes in the choice of clustering as more and more pentamers are added. Phosphorylation and acetylation of VP1 occur in virus-infected cells, and these modifications could further regulate assembly³⁰. Indeed, in polyoma, phosphorylation is required for virion stability³¹. Mistakes in assembly could also occur if C-terminal arms folded back into their pentamer of origin. Inspection of Fig. 3 shows that if the C helix is unwound, an arm can easily insert into its own subunit, forming contacts identical to those made in an intact virus particle by an arm from an adjacent pentamer and so disrupting assembly. It is therefore critical to maintain the arms exposed even under conditions that favour association. Accurate assembly in the nucleus of an infected cell may require participation of a 'chaperone' protein³² to prevent or to reverse the folding back of C-terminal arms, or to inhibit premature assembly in the cytoplasm.

In the icosahedral plant and insect RNA viruses with 180 subunits, such as TBSV, there are also alternative bonding modes³³. But because their design conforms to a so-called 'quasi-equivalent' arrangement³⁴, similar contacts can be made by all

subunits. The building block is a dimer, and it has two distinct conformations in the virion³⁵. Assembly requires a selection between two states of an extended interface, related by a hinge-like motion. The choice is determined by the folding or unfolding of an N-terminal arm. Propagated formation of an internal framework, composed of 60 interlinked arms, seems to direct the assembly process³⁶.

The SV40 structure shows that VP1 pentamers are linked primarily by C-terminal arms inserted into the folded structure of their neighbours, with potential additional contributions from divalent cations and disulphide bonds. The 'tied-together' character of the SV40 shell presents one general solution to the problem of how to generate a structure with several kinds of contacts among standard building blocks. We suggest that sub-cellular assemblies frequently will be found to exhibit these sorts of linkages. Many such assemblies seem to be flexible yet highly specific. Interaction through arms can ensure specificity without requiring a rigid geometry and without imposing strong restrictions on symmetry. □

Received 6 September; accepted 9 October 1991.

1. Tooze, J. *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1980).
2. Rossman, M. G. & Johnson, J. E. *A. Rev. Biochem.* **58**, 533-573 (1989).
3. Harrison, S. C. *Harvey Lect.* **85**, 127-152 (1991).
4. Klug, A. *J. molec. Biol.* **11**, 424-431 (1965).
5. Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. *Nature* **295**, 110-115 (1982).
6. Salunke, D., Caspar, D. L. D. & Garcea, R. L. *Cell* **46**, 895-904 (1986).
7. Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. *Acta crystallogr.* **18**, 169-179 (1965).
8. Argos, P. & Rossmann, M. *Acta crystallogr.* **B32**, 2975-2979 (1976).
9. Terwilliger, T. & Eisenberg, D. *Acta crystallogr.* **A39**, 813-817 (1983).
10. Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **16**, 39-45 (1963).
11. Bricogne, G. *Acta crystallogr.* **A32**, 832-847 (1976).
12. Olson, A. J., Bricogne, G. & Harrison, S. C. *J. molec. Biol.* **171**, 61-93 (1983).
13. Rossmann, M. G. *et al. Nature* **317**, 145-153 (1985).
14. Moreland, R. B. & Garcea, R. L. *Virology* (in the press).
15. Moreland, R. B., Montross, L. & Garcea, R. L. *J. Virol.* **65**, 1168-1176 (1991).
16. Brady, J. N., Winston, V. D. & Consigli, R. A. *Virology* **23**, 717-724 (1977).
17. Christiansen, G., Landers, T., Griffith, J. & Berg, P. *J. Virol.* **21**, 1079-1084 (1977).
18. Walter, G. & Deppert, W. *Cold Spring Harb. Symp. quant. Biol.* **39**, 255-257 (1974).
19. Blasquez, V., Beecher, S. & Bina, M. *J. biol. Chem.* **258**, 8477-8484 (1983).
20. Baker, T. S., Drak, J. & Bina, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 422-426 (1988).
21. Fried, H., Cahan, L. D. & Paulson, J. C. *Virology* **109**, 188-192 (1981).
22. Freund, R., Calderone, A., Dawe, C. J. & Benjamin, T. L. *J. Virol.* **65**, 335-341 (1991).
23. Dubensky, T. W., Freund, R., Dawe, C. J. & Benjamin, T. L. *J. Virol.* **65**, 342-349 (1991).

24. Freund, R., Garcea, R. L., Sahli, R. & Benjamin, T. L. *J. Virol.* **65**, 350-355 (1991).
25. Leavitt, A. D., Roberts, T. M. & Garcea, R. L. *J. biol. Chem.* **260**, 12803-12809 (1985).
26. Salunke, D., Caspar, D. L. D. & Garcea, R. L. *Biophys. J.* **56**, 887-900 (1989).
27. Garcea, R. L., Salunke, D. & Caspar, D. L. D. *Nature* **329**, 86-87 (1987).
28. Kiselev, N. A. & Klug, A. *J. molec. Biol.* **40**, 155-171 (1969).
29. Baker, T. S., Caspar, D. L. D. & Murakami, W. T. *Nature* **303**, 446-448 (1983).
30. O'Farrell, P. Z. & Goodman, H. M. *Cell* **9**, 289-298 (1976).
31. Garcea, R. L. & Benjamin, T. L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3613-3617 (1983).
32. Ellis, R. J. & van der Vries, S. M. A. *Rev. Biochem.* **60**, 321-348 (1991).
33. Harrison, S. C. *Biophys. J.* **32**, 139-153 (1980).
34. Caspar, D. L. D. & Klug, A. *Cold Spring Harb. Symp. quant. Biol.* **27**, 1-24 (1962).
35. Harrison, S. C. *Trends Biochem. Sci.* **9**, 345-351 (1984).
36. Sorger, P. K., Stockley, P. G. & Harrison, S. C. *J. molec. Biol.* **191**, 639-658 (1986).
37. Lattman, E. E. *Science* **208**, 1048-1050 (1980).
38. Crawford, J. thesis, Harvard Univ. (1977).
39. Harrison, S. C. *et al. Meth. Enzym.* **114**, 211-237 (1985).
40. Filman, D. J. *et al. EMBO J.* **8**, 1567-1579 (1989).
41. Arnold, E. & Rossmann, M. G. *J. molec. Biol.* **211**, 763-801 (1990).
42. Hogle, J. M., Kirchhausen, T. & Harrison, S. C. *J. molec. Biol.* **171**, 95-100 (1983).

ACKNOWLEDGEMENTS. We thank D. L. D. Caspar and R. Garcea for comments on the manuscript, S. Gamblin for discussion and assistance, K. Schwarz for computing, R. Hellmiss for illustrations, the SERC for beamtime on stations 9.6 and 7.2 at Daresbury, and CHES for time on beamline A1, and the staff at both of these facilities for their assistance. This research has been supported by NIH grants to S.C.H. and T.L.B. and by the Howard Hughes Medical Institute.

LETTERS TO NATURE

Ca II emission from old red giants in the globular cluster M15

A. K. Dupree & B. A. Whitney

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

MURPHY *et al.*¹ recently observed characteristic emission of Ca II from the globular cluster M15, and argued that this emission came from a population of primordial binary stars. Their conclusion appears theoretically attractive in relation to the study of the dynamics and evolution of globular clusters^{2,3}. We argue here, however, that the Ca II K-line emission in various types of hard binary systems differs in strength and width from the M15 spectra, and we demonstrate that the M15 emission closely resembles that from metal-deficient red giants of the halo population in our Galaxy. Chromospheric activity, indicated by emission in Ca II, H α and Mg II, occurs in red giants in both globular clusters and halo-population field giants⁴⁻⁶. A simple detection of Ca II emission is thus not a unique signature of binaries, and in the case of M15 we argue that red giants are the more likely cause.

Objects in the core of M15 are of great interest because theoretical studies suggest that dynamical evolution of a globular cluster affects its stellar contents. Murphy *et al.*¹ obtained spectra

TABLE 1 Ca K-line profiles

Target	V	(B-V) ₀	Width* (Å)	Relative flux (Ca K/flux at 3,930 Å)	Ref.
'Core' (AC214?)	14.75	0.10	1.1	0.31	1
AC623 (M15)	13.50	1.04	1.2	0.47	1
α Boo (HD124897)	-0.04	1.23	1.4	0.67	†
HD6833	6.75	1.00	1.2	0.57	†
HD187111	7.74	1.13	1.3	0.42	†
W UMa (HD83950)	7.8	0.4-0.5	0.7	1.2	†
FK Comae (HD117555)	8.18	0.88	5	>1	†
α^2 CrB (HD146361)	5.7	0.5	2.7	0.7	11
λ And (HD222107)	3.88	1.01	1.4-1.8	5-10	10

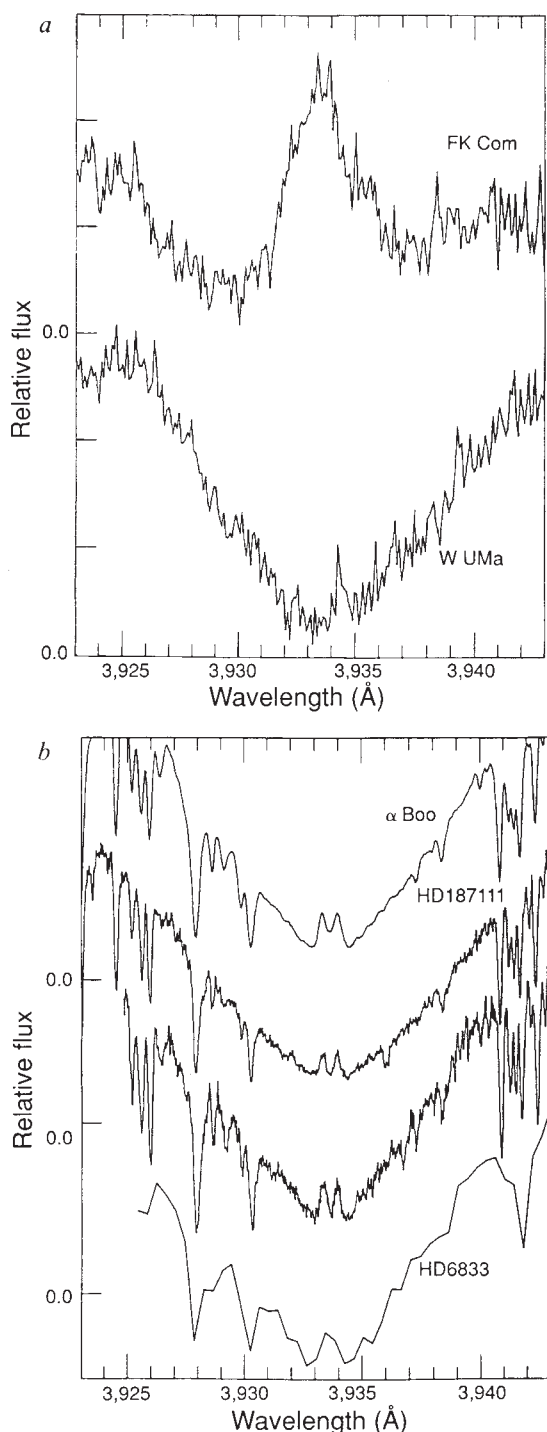
* Full width at base of K-line emission core [$\Delta\lambda(K_1)$], corrected for instrumental resolution.

† This paper.

at three positions in and near the core of M15. All of the spectra revealed strong absorption in the Ca II resonance doublet at 3,968.5 Å (H-line) and 3,933.7 Å (K-line), which is characteristic of cool stars. Two of the M15 objects identified^{1,7} as AC214 and AC623 show chromospheric emission centred in the deep Ca absorption core. Because rapidly rotating cool stars can produce strong Ca II emission, the emission from M15 was attributed¹ to hard binary systems ($a \approx 0.4-0.5$ AU), with enhanced chromospheric magnetic activity. Hard binaries have internal

orbital velocities larger than the local stellar velocity dispersion, and tend not to be disrupted by stellar encounters in the dense core of a globular cluster.

We have obtained Ca II K-line spectra of tightly bound binary systems and cool giant stars. In addition, spectra of RS CVn binaries are in the published literature. These stars and stellar systems are all reasonable candidates for objects that may be present in M15. They have orbital velocities (if binary) that in most cases exceed the observed⁸ velocity dispersion of $\sim 25 \text{ km s}^{-1}$ in the cusp of M15, or $14.2 \pm 1.9 \text{ km s}^{-1}$ within 20 arcsec of the core. Eleven stars are considered: a contact binary (W UMa); a rapidly rotating coalesced star (FK Comae); two RS CVn stars (σ^2 CrB and λ And); binaries with separations $a \approx 0.4\text{--}0.5 \text{ AU}$; a cataclysmic variable (YY Dra); and three single giant stars (α Boo, HD6833 and HD187111).



Our spectra were obtained at the 1.5-m Tillinghast reflector and echelle spectrograph of the F. L. Whipple Observatory at Mount Hopkins, Arizona. The detector was either a dual-array Reticon ($2 \times 1,032$) or a Ford CCD ($2,048 \times 2,048$). Spectra were reduced with IRAF using Th-Ar lamp exposures bracketing the stellar exposures to establish the wavelength scale; the spectral resolution is $\sim 0.15 \text{ Å}$.

Two convenient parameters for characterizing the strength of the Ca II feature are the width of the emission at its base [$\Delta\lambda(K_1)$], and the strength of the emission core measured relative to the continuum $\sim 3 \text{ Å}$ short of the Ca II K-line. The ratio does not uniquely characterize the spectra, but it minimizes the effects of metal abundance variations and different instrumental configurations. We have simply assumed a constant effective background (arising from scattered light in the spectrograph or the sky, and/or low spectral resolution) amounting to the value at the deepest part of the absorption profile, and measured the Ca K-emission and the continuum relative to this background. Table 1 gives parameters for various objects.

Hard binaries might be primordial or formed by stellar encounters⁹ in the core of M15. The semimajor axes of these systems must be $< 0.5 \text{ AU}$, which leads to a rotational synchronization of the stars with the orbital motion. Binaries of the W UMa type are a useful example because these contact systems have components of low-mass stars. W UMa itself has a period of 0.336 d, and velocity of each component, $v \sin i \approx 150 \text{ km s}^{-1}$. Such velocities substantially broaden spectral features, except for possible variable emission that may be associated with one or another component (see Fig. 1a). This spectrum contrasts with the narrow photospheric absorption lines that have been observed in M15.

RS CVn systems, such as λ And (G8 III-IV), with $a \sin i = 0.14 \text{ AU}$ qualify as hard binaries, but the Ca II emission is broader, and stronger by an order of magnitude¹⁰ than observed in M15. The λ And system is not synchronized and the orbital period (20.5 d) is shorter than the photometric period (54.2 d). A synchronous RS CVn binary is represented by σ^2 CrB with wide, strong Ca II emission and spectral features broadened by its $\sim 25 \text{ km s}^{-1}$ rotational velocity¹¹.

Another possible identification is a coalesced binary such as FK Comae. These objects are extremely rapidly rotating giant stars ($v \sin i = 154 \pm 4 \text{ km s}^{-1}$)¹², which are believed to result from the merging of two single stars¹³ as would occur during the evolution of a W UMa-type system. Rapid rotation again smears out the spectral features. Our observations show (Fig. 1a) that the emission is much stronger than observed in the M15 objects, and much wider because of the rotational broadening.

Cataclysmic variables might be another candidate, as they are close binaries composed of a degenerate primary star surrounded by an accretion disk, and a cool secondary component.

FIG. 1 Echelle spectra in the region of the Ca II K-line. a, Two rapidly rotating objects: FK Comae (thought to be a coalesced binary star) and W UMa (the prototypical contact binary system). Emission in the W UMa spectrum, seen here at $3,934.3 \text{ Å}$, varies in wavelength with orbital phase. b, Three red giants: α Boo, HD187111 wavelength and HD6833. The wavelength scale has been adjusted to align the Ca II K-line emission at $3,933.7 \text{ Å}$. All three stars show emission that is comparable in appearance to the emission detected in M15. The lower curve repeats the spectrum of HD6833 summed to a resolution of 0.4 Å which was the resolution used¹ in the detection of Ca II emission in the core of M15. Emission is still present at the lower resolution, but the central reversal in the line profile cannot be detected.

High-resolution spectroscopy of the Ca II H- and K-lines has not been obtained as far as we know, but time-resolved spectroscopic studies of the near-infrared Ca II lines ($\lambda 8498$ and $\lambda 8542$) in the cataclysmic variable YY Dra¹⁴ indicate what might be expected from the H- and K-lines. YY Dra shows broad emission lines ($1,300 \text{ km s}^{-1}$) on which are superposed variable narrow (full width at half maximum $60\text{--}110 \text{ km s}^{-1}$) emission features. The narrow component is visible only during half of the orbital period, and results from reprocessed radiation in the atmosphere of the cool star facing the white dwarf. Substantial velocity variation (200 km s^{-1}) of the narrow component occurs during the half-period (1.98 h) when it is visible. During a 40-min exposure on YY Dra (equivalent to the exposure on the M15 objects) at the phases when the narrow feature is visible, the emission will move by $\sim 1 \text{ \AA}$ as its equivalent width changes by a factor of 3. A rather fortuitous set of circumstances is required to capture the narrow emission from a cataclysmic variable in the core of M15. Moreover, there is no indication of a broad emission feature from an accretion disk.

Binary systems containing one giant star ($M_{\text{total}} \approx 2M_{\odot}$), with separation $a \approx 0.4\text{--}0.5 \text{ AU}$, have orbital periods of $65\text{--}95 \text{ d}$. A Ca II survey¹¹ contains three systems with giant stars as components, and periods of $75\text{--}85 \text{ d}$: HD32357, HD7672 and HD9746. All show emission with a strength comparable to, or greater than, the continuum, and the width of the Ca II K-emission exceeds by a factor of 2–3 (for the two stars measured) that observed in M15. Thus these spectra also differ substantially from the M15 objects.

Single stars chosen for comparison include α Boo, a slightly metal-deficient field K giant star, whose Ca K-emission parameters are commensurate with the M15 objects. The remaining two stars, HD6833 and HD187111, are the most relevant, because these metal-deficient field giants¹⁵ belong to the halo population of our Galaxy, have metal abundances and kinematics similar to those of the globular clusters, and can be assigned an age of 15 Gyr, which is comparable to that of M15. Such field stars show strong chromospheric Mg II emission⁵, Ca II and H α line profiles, and Mg II surface fluxes similar to those found in red giants in the globular cluster NGC6752⁶. The profiles displayed in Fig. 1b are remarkably similar in width and in relative flux to the profiles obtained from the stars in M15. Although of lower quality, profiles of two red giants in the globular cluster NGC6752 have similar parameters⁶. Figure 1b also shows a profile at echelle resolution that is binned over $0.4 \text{ \AA}$ to correspond to the resolution of the M15 objects. The structure of a centrally reversed line is not observed at this resolution. The profile seems comparable to the M15 spectra, which contain notches in the emission cores that cannot be confidently distinguished from the noise level. Thus metal-deficient red giants adequately match the M15 profiles.

To explain chromospheric emission from a population of old single stars in which magnetic activity has decayed, the spectroscopic evidence suggests that a hydrodynamic chromosphere occurs, perhaps driven by photospheric pulsations. This would account⁴ for the variable asymmetric H α profiles and the observed velocity variations. These spectroscopic criteria suggest that hydrodynamic chromospheres occur in stars brighter than $M_V = -1.25$, or ~ 2 magnitudes below the tip of the red giant branch in M15. In addition, Ca II K-emission can be found in

field giants as faint as $M_V = 1.5$ (A.K.D. and G. H. Smith, in preparation) corresponding to at least $V \approx 16.9$, or 4.5 magnitudes below the tip of the giant branch in M15.

Clearly Ca II emission is not a unique signature of the binary nature of the objects in M15, because we find it hard to identify a known binary system with emission similar to that in M15. In addition, the narrow photospheric absorption lines in the M15 spectra rule out the most rapidly rotating systems.

Our conclusion that red giants produce this emission is further supported by ultraviolet and optical images of the M15 core region. Objects in the central cusp have magnitudes and colours consistent with a clump of red giants 2–3 magnitudes below the tip of the giant branch⁸. Moreover, a spectrum covering wavelengths $5,160\text{--}5,210 \text{ \AA}$ at the centre of M15 resembles⁸ those of red giants distant from the core. International Ultraviolet Explorer spectra of a $10 \times 20 \text{ arcsec}$ region of the core of M15 also show a typical red giant spectrum at long wavelengths ($2,000\text{--}3,000 \text{ \AA}$), whereas the short-wavelength spectrum ($1,200\text{--}2,000 \text{ \AA}$) resembles horizontal branch stars¹⁶. No emission features or P Cygni profiles, characteristics of hard binaries, are present in the short-wavelength spectrum.

The recent Hubble Space Telescope WF/PC image (U-band) of the core of M15 shows¹⁷ 21 objects within the $3 \times 1 \text{ arcsec}$ aperture used for the M15 observations. Two of the brightest objects are candidates to produce the observed spectrum: AC214 and AC215. Photometry in U-, V- and I-bands from the WF/PC images suggests (B. Yanny, personal communication) that AC215 is definitely a red giant; AC214 is hotter than AC215, but it is not known from preliminary reductions whether AC214 lies on the asymptotic giant branch or the red giant branch, or is associated with, or lies above the horizontal branch. Inspection of the M15 spectrum¹ perpendicular to the dispersion shows (B. W. Murphy, personal communication) that the emission is probably associated with AC214. A recent spectrum of the core of M15, obtained by Murphy with the spectrograph slit perpendicular to its earlier orientation, and excluding AC215, appears similar to the published spectrum.

Further observations of the M15 core should seek variability in the line flux which may result from one or from a combination of several of the following phenomena: rotation of magnetic activity on the surface of a star, periodic pulsation enhancing the atmospheric density and temperature in a hydrodynamic atmosphere, or hot regions resulting from reprocessed radiation in a cataclysmic variable or contact binary. The timescale of flux variation may indicate its source. The best discriminator would be radial velocity variations of the emission. The photospheric absorption lines visible in the M15 spectra suggest stellar rotation velocities $\leq 50 \text{ km s}^{-1}$, and velocity excursions should be detectable if they exist. A high-resolution spectrum of a single object could reveal the central reversal (K_3) of the Ca II core which appears in giant stars, but is absent in most binary systems.

If single red giants dominate the stellar population in the core, this appears to confront current interpretations of M15 as a post-core-collapse globular cluster where mass segregation has occurred, and where the core binary density is expected to be 5–10 times higher than single stars³. In this context, the recent high-resolution images of the centre of M15 taken with the Hubble Space Telescope planetary camera¹⁷ show a large core, and a light distribution in agreement with either pre-collapse or post-collapse models. $\square$

Received 29 May; accepted 30 September 1991.

1. Murphy, B. W. et al. *Nature* **351**, 130–132 (1991).
2. Elson, R. Hut, P. & Inagaki, S. A. *Rev. Astr. Astrophys.* **25**, 565–601 (1987).
3. Gao, B., Goodman, J., Cohn, H. & Murphy, B. *Astrophys. J.* **370**, 567–582 (1991).
4. Smith, G. H. & Dupree, A. K. *Astr. J.* **95**, 1547–1557 (1988).
5. Dupree, A. K., Hartmann, L. & Smith, G. H. *Astrophys. J.* **353**, 623–635 (1990a).
6. Dupree, A. K. et al. *Astrophys. J.* **361**, L9–L13 (1990b).
7. Aurière, M. & Cordoni, J.-P. *Astr. Astrophys. Suppl. Ser.* **46**, 347–354 (1981).
8. Peterson, R. C., Seitzer, P. & Cudworth, K. M. *Astrophys. J.* **347**, 251–265 (1989).
9. Davies, M. B., Benz, W. & Hills, J. G. *Astrophys. J.* **381** (1991).

10. Baliunas, S. L. & Dupree, A. K. *Astrophys. J.* **252**, 668–680 (1982).
11. Strassmeier, K. G., Fekel, F. C., Bopp, B. W., Dempsey, R. C. & Henry, G. W. *Astrophys. J. Suppl.* **72**, 191–230 (1990).
12. Rucinski, S. M. *Publ. Astr. Soc. Pac.* **102**, 306–311 (1990).
13. Bopp, B. W. & Stencel, R. E. *Astrophys. J.* **247**, L131–L134 (1981).
14. Mateo, M., Szkody, P. & Garnavich, P. *Astrophys. J.* **370**, 370–383 (1991).
15. Bond, H. E. *Astrophys. J. Suppl.* **44**, 517–533 (1980).
16. Dupree, A. K. et al. *Astrophys. J.* **230**, L89–L92 (1979).
17. Lauer, T. R. et al. *Astrophys. J.* **369**, L45–L49 (1991).

Near miss of the Earth by a small asteroid

J. V. Scotti*, D. L. Rabinowitz* & B. G. Marsden†

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

THE 0.91-m Spacewatch Telescope, on Kitt Peak in Arizona, is being used to search for new Earth-approaching asteroids. On the night of 18 January 1991 a streaked image extending over more than 161 arcseconds was recorded. Further observations over the next 4.6 hours showed that the images were of an object travelling in an independent orbit around the Sun; it was given the asteroidal designation 1991 BA. Orbital computations, presented here, indicate that 1991 BA was only 0.0053 AU from the Earth at the time of discovery, closing to 0.0033 AU at the last detection. Extrapolation of the calculated orbit shows that the object passed only 0.0011 AU (170,000 km) from the Earth 12 hours after it was found. The observed brightness translates into an absolute visual magnitude 28.9, corresponding to a diameter of only 5–10 m. 1991 BA is the closest and smallest asteroid yet observed outside the Earth's atmosphere.

Our instrumentation and observing technique have been described elsewhere^{1–3}. Briefly, we use a Tektronix TK2048 CCD located at the newtonian focus in scanning mode. With the telescope drive turned off, the accumulating charges in the CCD are read out so that the rate of line transfers matches the rate of the sky drift. At the same time, a workstation computer at the telescope site searches for the trailed images of asteroids near the earth. More distant objects, such as asteroids in the main belt, are also detected by their consistent motion over three consecutive scans. The computer also displays the CCD image and allows the observer to indicate trailed images that the software has not flagged on its own.

Our goal is to survey in detail the population of Earth-approaching asteroids in order to study the physical properties of the objects in near-Earth space, to discover objects dangerous to humanity and to identify easily accessible candidates in the near-Earth population as possible exploration targets. In the first 10 months of full-time surveying with our automatic detection system, beginning in September 1990, we have detected 15 new Earth-approaching asteroids, as well as a monthly average of ~2,000 other asteroids^{4,5}.

1991 BA was discovered when Rabinowitz noticed a faint trail ~161 arcsec in length while monitoring the computer display and immediately measured the pixel location of the trail's endpoints. We estimated the object's rates of motion in right ascension and declination (although the direction of motion was not yet known) and used a linear extrapolation to estimate the trajectory across the sky. After the object was recovered in a scan an hour later and again an hour after that, future positions were estimated from an improved linear extrapolation, and the object was reobserved at intervals ~15–30 minutes. After the third observation, it became apparent that there was some angular acceleration. Not only had the estimated position fallen short of the observation position, but the trail had lengthened considerably.

We obtained a total of nine images between 05:25:43 UT and 10:03:33 UT. Eight were usable for making astrometric measurements. Because of its rapid angular motion, the object had a peak brightness of only ~1 σ above the noise in the surrounding image. Its magnitude per pixel across the seeing profile was $V \approx 23$, and the apparent magnitude integrated over the length of the trail varied from $V = 18.2$ to $V = 17.2$. Figure 1 shows eight images of the object from its discovery to its last observed image. The object's faint trail is just visible near the centre of each image.

TABLE 1 Positions and brightness of 1991 BA

Date (UT)	RA h min s	Dec ° ' "	V	H
1991 01 18.22620	08 05 09.6	+12 35 49	18.2	29.0
1991 01 18.30358	08 12 29.8	+11 29 05	18.0	29.1
1991 01 18.31493	08 13 46.8	+11 17 03	17.9	29.1
1991 01 18.32672	08 15 10.9	+11 03 50	17.7	28.9
1991 01 18.35716	08 19 11.0	+10 25 46	17.5	28.8
1991 01 18.36578	08 20 25.8	+10 13 48	17.5	28.9
1991 01 18.40058	08 26 05.2	+09 19 08	17.4	28.9
1991 01 18.41914	08 29 34.4	+08 45 10	17.2	28.8

Dates are UT, positions are astrometric referred to the Equator and equinox of B1950.0. V is the apparent magnitude. H is the absolute magnitude assuming a slope factor $G=0.15$.

The astrometric reductions for Spacewatch scanning have been described elsewhere⁶. The reference catalogue used is the AGK3, which for the Northern Hemisphere gives better proper-motion estimates at the epoch of observation and smaller residuals than the SAO catalogue that we used in the past. The field of view with the TK2048 is ~38 arcmin, allowing for defects, compared with only ~9 arcmin with our previous 320 × 512 pixel CCD. Thus many more reference stars become available in the same amount of time. Unfortunately, the $f/5$ newtonian system results in significant coma in the images as they drift on and off the edges of the CCD. The effects of coma are most apparent in the ends of streaked images. As the CCD is nearly centred on the optical axis, the coma is nearly symmetric. Because of the faintness of the trail and the coma, the endpoints of the streak are rather ill-defined. After the best possible measurements were made of the endpoints of each streak, these positions were averaged to give as consistent a measure as possible of the midpoint of the streak. Because each exposure had a duration of only ~3 min, it was reasonable to ignore the acceleration of the object's apparent motion during the time of exposure. The observation time to be recorded is the average of the start and end times of the trail's exposure, determined from its endpoints, and the integration time is the difference of the end time and the start time. The astrometric reduction of the measurements corresponding to these mid-exposures, referred to the mean equator and equinox of 1950.0, is shown in Table 1.

Once the astrometric positions were available, we established that they were consistent with a heliocentric orbit (see below), and the object was given the asteroid designation 1991 BA (ref. 7). The orbit, whose elements are shown in Table 2 (ref. 8), satisfied all but the seventh of the eight positions within 1 arcsec; the seventh yielded a residual of ~4 arcsec. This solution indicates that the object's distance from the centre of the Earth decreased from 0.00527 AU at the first observation to 0.00334 AU at the eighth with a minimum value of 0.00114 AU (less than half the distance to the Moon) around 17:19 UT. The uncertainty

TABLE 2 The orbit of 1991 BA

a	e	i	Ω	ω	T (UT)
2.243	0.682	1.96°	118.34°	70.58°	1991 March 2.062
Change in the orbit					
Δa	Δe	Δi	$\Delta \Omega$	$\Delta \omega$	ΔT (days)
-0.0783	-0.0086	+0.06°	-0.06°	+1.35°	+0.596

a is the semimajor axis in AU, e is the orbital eccentricity, i is the orbital inclination in degrees, Ω is the longitude of the ascending node in degrees, ω is the argument of perihelion in degrees and T is the time of perihelion passage. The epoch of the orbit is near the middle of the observation interval, and the angular elements are referred to the ecliptic and mean equinox of B1950.0. The changes in the elements are over a 48-hour interval centred on the time of closest approach to Earth.

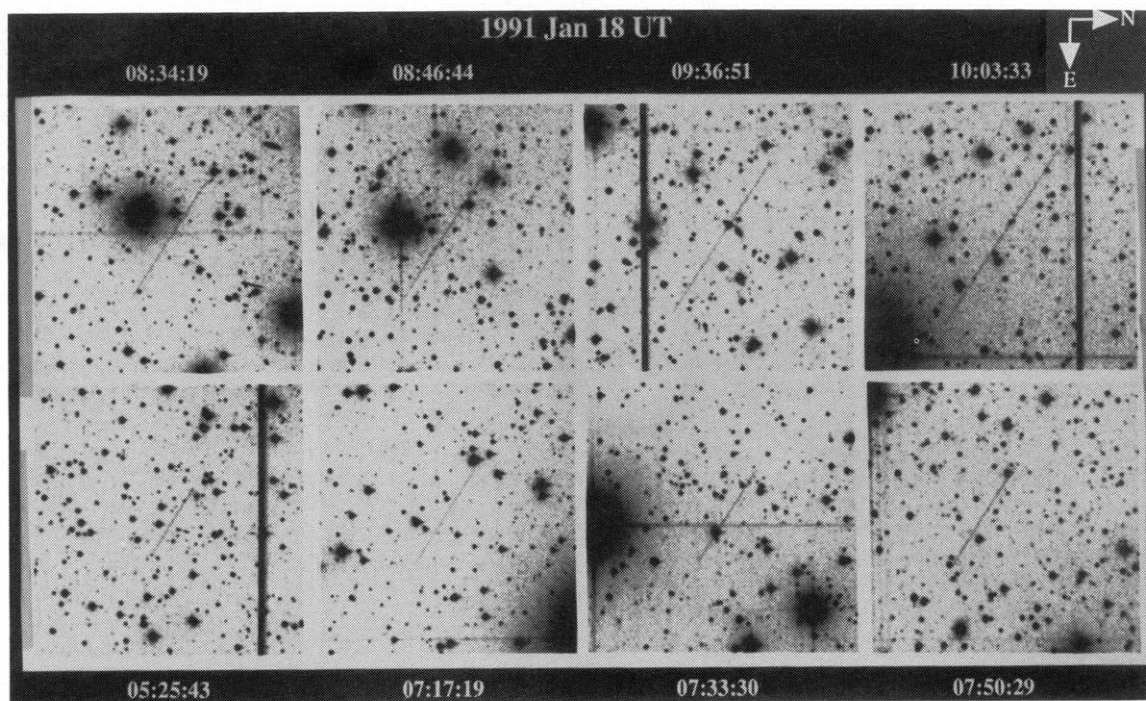


FIG. 1 Eight images of asteroid 1991 BA obtained on the night of 18 January 1991 with the Spacewatch Telescope. The mid-times of exposure are displayed next to the individual images. Each exposure is 2.8 min. The asteroid is the faint trail near the middle of each frame. The trail length

increased from 161 arcsec up to 457 arcsec during the observation interval. The image scale is 1.21 arcsec per pixel and each image is 10.3 arcmin on a side. Note that the sequence of images runs from bottom left to right, then top left to right.

in this solution can best be estimated by eliminating the eighth observation and retaining the seventh; this fit is almost as good, but the residual of the eighth observation then increases to 7 arcsec. This alternative solution gives values for the above three geocentric distances that are 0.8% larger and increases the object's semimajor axis by $\sim 2\%$. We have also considered the possibility that 1991 BA is in orbit about the Earth. Thanks to the large motion of the Spacewatch Telescope about the centre of the Earth during the interval covered by the observations, parallax could be used to show that no meaningful geocentric orbit was possible; indeed, the closest fits to the observations required near-infinite orbital eccentricity, and the best elliptical solution, which turned out to be almost parabolic, gives systematic residuals in excess of 6 arcmin. Although the uncertainty in the orbit is relatively large, it is of interest to examine the differential perturbations exerted by the Earth-Moon system on the object during its passage. Table 2 also shows the change in the orbital elements during the 48-hour interval surrounding that time.

The measured trail length of 1991 BA varied from ~ 161 to 457 arcsec. Because of background star images, it proved impossible simply to integrate all of the light contributing to the trail. We have therefore integrated across the width of the seeing disk perpendicular to the trail. The total brightness is the mean trail brightness, determined from a segment of the trail free from background stars and CCD defects, and scaled to the length of the trail. From this measurement and a magnitude calibration determined from the observations of stars that have colours close to those of the Sun in photometric standard region Selected Area 57, we determined the visual magnitudes (V) listed in Table 1. Using the geometry of the orbit in relation to the Earth, we derived the absolute magnitudes (H) listed in Table 1 by the method outlined by *Bowell et al.*⁹ The mean absolute magnitude H is 28.9 ± 0.4 , which includes the random error of the brightness measurements (± 0.2) and the systematic error from the magnitude calibration (± 0.3). Assuming a mean albedo ranging from 0.048 to 0.186, which is the range spanned by most asteroids, the mean diameter of 1991 BA falls in the range of 10.2 and 5.2 m.

Is 1991 BA an asteroid or a meteoroid? That is the question that demonstrates the importance of the find. Other objects of comparable or smaller size have been observed near the Earth, but only after they have entered the atmosphere and appeared as fireballs. One such object with an estimated diameter of ~ 4 m was visible over the western United States in 1972 (ref. 10). Another, estimated to be ~ 200 tons before its entry into the atmosphere, produced the Sikhote-Alin fall in Siberia in 1947. The object shattered at high altitude and produced ~ 200 craters and holes ranging in size from less than a metre up to 26.5 m (ref. 11). Larger objects with diameters ≥ 100 m have been observed by us and others at distances usually greater than 0.01 AU from the Earth and recognized as Earth-approaching asteroids. The smallest of these known asteroids, other than 1991 BA, is 1990 UN, which we discovered at a distance of 0.12 AU from the Earth in October 1990. It has a diameter of 60–120 m, assuming the same albedo range as for 1991 BA, and an upper limit of 200 m as determined by Doppler-shift radar (S. Ostro, personal communication). 1991 BA bridges the observational gap between the smallest asteroids and the largest meteoroids. It is the only observed meteoroid with a measured trajectory and diameter unaltered by passage through the atmosphere. It is the only observed asteroid in the size range of objects we can expect to strike the Earth at least once in a person's lifetime. The only natural objects observed outside the Earth's atmosphere of comparable or smaller sizes are the nuclei of several recently observed Sun-grazing comets of the Kreutz family, with diameters estimated indirectly from their sublimation rates¹². Perhaps 1991 BA itself may have originated as a cometary fragment. Further observations of these small asteroids will allow us to determine their size-frequency distribution and hence the relative contribution of cometary and asteroidal material to the near-Earth environment⁵.

Note added in proof: Four new asteroidal objects with diameters less than 30 m were discovered at distances less than 0.03 AU from Earth with the Spacewatch Telescope in October and November 1991: 1991 TT, 1991 TU and 1991 VA were discovered in the same manner as 1991 BA—by visually detecting their long trailed images. 1991 VG was found by the automatic

software as a stellar image with rates under one degree per day. □

Received 31 July; accepted 15 October 1991.

1. Gehrels, T. *Space Sci. Rev.* **58**, (1991).
2. Gehrels, T., McMillan, R. S., Scotti, J. V. & Perry, M. L. in *CCDs in Astronomy Astr. Soc. Pacific Conf. Series*, Vol. 8 (ed. Jacoby, G.) 51–52 (Provo, 1990).
3. Rabinowitz, D. L. *Astr. J.* **101**, 1518–1529 (1991).
4. Scotti, J. V., Gehrels, T. & Rabinowitz, D. L. in *Asteroids, Comets, and Meteors*, 1991 (Lunar and Planetary Institute, in the press).
5. Rabinowitz, D. L. in *Asteroids, Comets, and Meteors*, 1991 (Lunar and Planetary Institute, in the press).
6. Gehrels, T., Marsden, B. G., McMillan, R. S. & Scotti, J. V. *Astr. J.* **91**, 1242–1243 (1986).
7. Marsden, B. G. *IAU Circ. No.* 5,172 (1991).
8. Marsden, B. G. *Minor Planet Circ. No.* 17,833 (1991).
9. Bowell, E. *et al.* in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 524–556 (University of Arizona Press, Tucson, 1989).
10. Jachia, L. G. *Sky Telesc.* **48**, 4–9 (1974).
11. Krinov, E. L. in *The Solar System, Vol IV, Meteorites and Craters* (eds Middlehurst, B. & Kuiper, G. P.) 208–234 (University of Chicago Press, 1963).
12. MacQueen, R. M. & St Cyr, O. C. *Icarus* **90**, 96–106 (1991).

ACKNOWLEDGEMENTS. We thank T. Gehrels and G. V. Williams for reviewing the manuscript and adding helpful comments and criticisms. We also thank R. S. McMillan, M. L. Perry and the many Spacewatch supporters.

Absence of a metallic phase at high pressures in C₆₀

M. Núñez Regueiro*†, P. Monceau*, A. Rassat‡, P. Bernier§ & A. Zahab§

* Centre de Recherches sur les Très Basses Températures, CNRS, BP 166X Cedex 9, 38042 Grenoble, France

‡ Laboratoire d'Activation Moléculaire, Ecole Normale Supérieure, 24 Rue Lhomond, 75231 Paris, France

§ Groupe de Dynamique des Phases Condensées, Université de Montpellier, 34060 Montpellier, France

THE bonding between molecules in bulk solid C₆₀ is extremely weak, making it a narrow-band semiconductor with an energy gap of 1.5 eV (ref. 1) for the face-centred cubic phase. Doping with alkali metals produces a metallic state which can be superconducting at temperatures as high as 33 K (ref. 2). The intermolecular coupling should have an important influence on the conductivity of the pure, semiconducting state: stronger coupling might induce a transition to a metallic or possibly even superconducting state, as is the case for silicon³, or it may result in a covalent solid such as diamond⁴. We have explored these possibilities by measuring the electrical resistivity of solid granular C₆₀ up to pressures of 25 GPa. Our results show that the magnitude of the gap and the resistivity decrease with increasing pressures as the sample volume⁵ decreases. But eventual gap closure to give a metallic state is not observed; instead, there is a sudden transition at 15–20 GPa to a more insulating phase, possibly with covalent intermolecular bonding.

Since C₆₀ became available in solid form⁶, there has been considerable interest in its properties. One of the most important properties is the high-temperature superconducting state that appears when it is metallized by intercalation and doping with alkali metals². Another way of rendering semiconductors metallic is by application of pressure. The increase of molecular wavefunction overlap due to the volume reduction thus caused can result in band-gap closure to a metallic state which can be superconducting, as in the case of silicon³. Conversely, we might expect the behaviour to resemble that of graphite, which at high pressure is transformed to a highly covalent solid (diamond) because of a change from sp² to sp³ hybridization of the carbon orbitals. Duclos *et al.*⁵ have observed a phase transformation in solid C₆₀ at pressures above 15 GPa. We have measured the electrical resistivity of C₆₀ under pressure, to determine the nature of its high-pressure state.

Soot containing different fullerenes was obtained in a pro-

cedure similar to that described by Haufler *et al.*⁷ A mixture of ~80% of C₆₀ and 20% of C₇₀ was extracted from the soot with boiling toluene. Carbon-60 was then separated from the mixture by column chromatography on neutral alumina followed by solvent drying. The purity of C₆₀ was checked by high-pressure liquid chromatography, infrared-, ultraviolet- and mass spectroscopy. No signal attributed to impurities was detected. The three samples, A, B and C, were pressed from the C₆₀ powder into 40-μm-thick platelets, which were then cut into rectangles (0.3 × 0.7 mm) and introduced into the 1-mm-diameter pressure cell. Sample C alone was heated in vacuum at 440 K for 2 hours before pressing to eliminate eventual traces of toluene.

The high-pressure measurements were performed using the sintered diamond anvil technique. We used a pyrophyllite gasket, through which platinum wires were passed to make the electrical contacts. We used a lead sample inside the cell as a manometer by monitoring its superconducting transition temperature. Soft steatite was used as the pressure medium in quasi-hydrostatic conditions. The pressure was applied at room temperature, and the measurement was performed in a helium cryostat with germanium and platinum resistors in good thermal contact with the anvils as thermometers. The sample resistance was 1–100 MΩ, and measured using four contacts with a Keithley 220 current source and a Keithley 181 nanovoltmeter. The width of the superconducting transition temperature was used to estimate the pressure gradient inside the cell, which was ~1 GPa. The resistivity of the samples was only measurable above ~5 GPa, because of both the quality and the high resistivity of the contacts at low pressures. The measurements were reproducible with cycling to low temperatures, but the value of the resistivity sometimes fluctuated at temperatures higher than 320 K.

We show in Fig. 1b the variation of the electrical resistivity for sample C at different pressures. There is a general tendency towards a decrease of the resistance of an order of magnitude per 10 GPa. We do not observe the perfect linear behaviour in the logarithmic inverse temperature plot that would correspond to a ideal semiconductor following the activation law $R = R_0 \exp(E_g/2k_B T)$, with a gap E_g proportional to the slope of the plotted curves. This variation makes it difficult to determine the evolution of the gap with pressure, although it is apparent that the average slope decreases with pressure up to the highest pressure, where it starts to increase. To quantify this observation we have calculated the derivative, which is shown in the insert of Fig. 1 as a function of temperature, and we choose its magnitude at 270 K as the value of the gap. Assuming that C₆₀ is an intrinsic semiconductor, we obtained for samples A, B and C the gaps shown in Fig. 2 as a function of pressure. We observe that for sample C the gap decreases monotonically with pressure up to 20 GPa. On the same plot, we show the value of the gap at ambient pressure extracted from the measurements of Mort *et al.*⁸ at temperatures below the ordering transition. The pressure dependence of the gap scales with the variation of volume with pressure, as determined by high-pressure X-ray measurements⁵. Thermal contraction of the lattice might therefore explain the temperature dependence of the gap. Measurements at ambient pressure on the simple cubic (s.c.) phase⁹, however, show a volume contraction of only 10% between room temperature and 11 K. The variation that we observe is much bigger and would correspond to a contraction of 30% down to 200 K. Another, probably more realistic, way of explaining the temperature variation of the gap is by the existence of disorder-induced localized states in the gap. Their variable-range hopping contribution to the electrical conduction should make the temperature dependence of the resistivity deviate considerably from the ideal behaviour. This interpretation is supported by Raman scattering measurements¹⁰ in which orientational disorder was observed at high pressure.

Above 20 GPa, the energy gap and resistance of sample C increase, implying that there is a transition to a more insulating phase. A similar effect is seen in the other samples, although at

† On leave of absence from Centro Atomico Bariloche, 8400 Bariloche, Argentina.

a lower pressure (15 GPa) for sample B. The C_{60} molecule is stable at these pressures³, and a low-symmetry phase has been detected⁵ at 15 GPa in nonhydrostatic conditions. This may be the case in sample B, as the pressure gradient was larger for this cell (1.5–2.0 GPa). We must also note that sample B was systematically cycled to 350 K, which may facilitate the transition, whereas samples A and C were only cycled up to room temperature. All the samples were left at 25 GPa for several days, after which they were totally insulating. The value of the gap above the critical pressure should then reflect the behaviour of a mixture of phases, one semiconducting and the other insulating, and not the high-pressure phase alone, which is obviously nonconducting and not measurable with our apparatus. We believe that our insulating high-pressure state corresponds to the low-symmetry phase observed in the X-ray experiment of ref. 5, and that pressures higher than 20 GPa can stabilize it even in quasihydrostatic conditions.

As our measurements scan only from 6 GPa upwards, the semiconducting gap that we have determined would correspond to the (s.c.) phase seen in refs 11 and 12 (note that the high-pressure X-ray data of ref. 5 did not distinguish between a

face-centred cubic (f.c.c.) and a s.c. structure). Its value, which by extrapolation would be ~ 0.5 eV at ambient pressure, is smaller than that obtained from band-structure calculations¹ (1.5 eV) and conductivity measurements at ambient pressure⁸ (1.9 eV) for the high-temperature f.c.c. phase, but corresponds correctly to the value extracted from the ambient pressure measurements of ref. 8 for the s.c. phase. The s.c. phase would then have a smaller gap than the f.c.c. This is unusual, as few materials develop structural phase transitions towards a more metallic phase as the temperature is lowered; to our knowledge, only Ag_2Te , HgI_2 and $(NbSe_4)_3I$ do so^{13–15}. The analogy with this last material is interesting, as the transition also involves a $q=0$ distortion wavevector, with a rearrangement of the selenium on the rings around the niobium chains. Optical measurements of the gap are necessary to clarify this issue, as determinations of the gap from resistivity can be affected by the disorder contribution.

From the few experiments performed so far we can draw the preliminary phase diagram for solid C_{60} shown in Fig. 3. Our results indicate that bringing together the C_{60} molecules by pressure reduces the energy gap of the s.c. phase, and this may

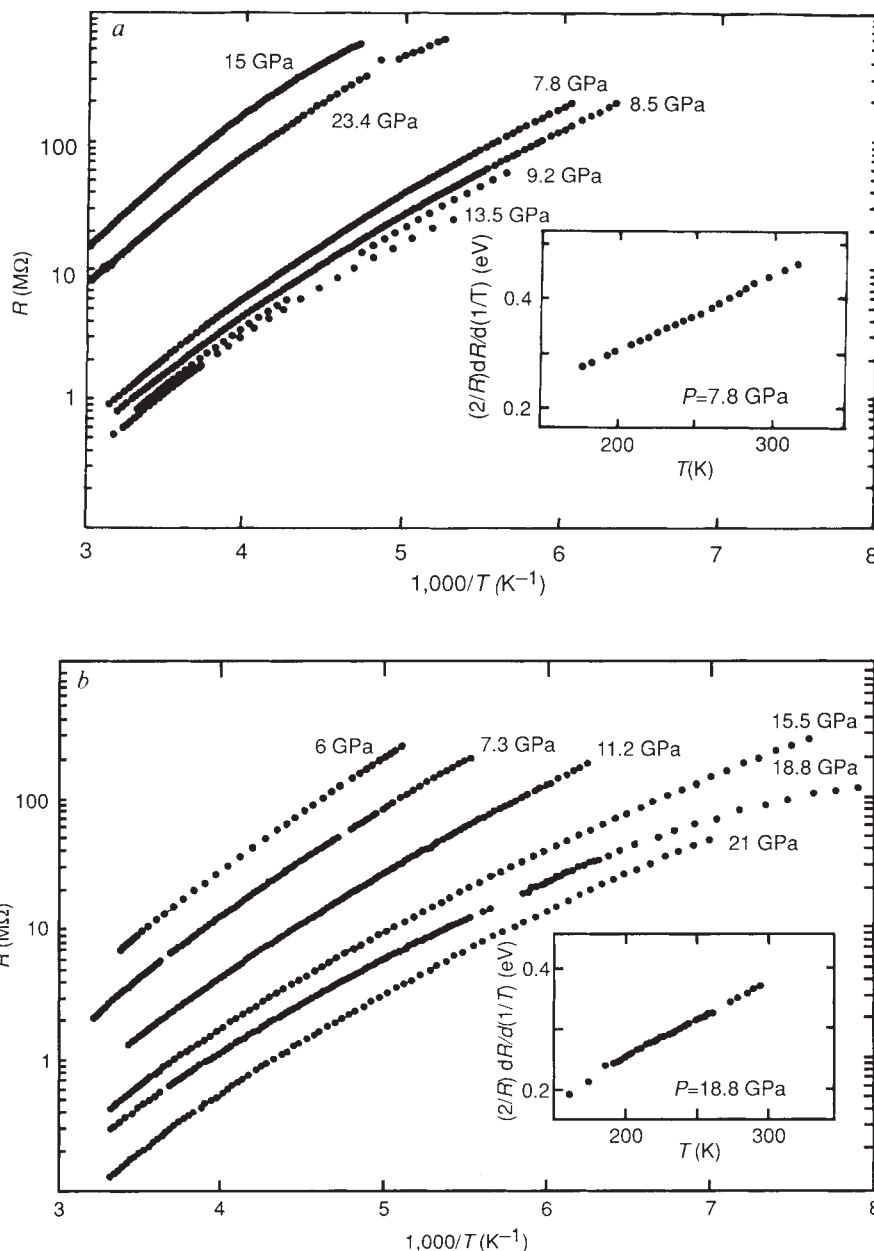


FIG. 1 *a*, Resistance curves for sample B as a function of inverse temperature for different pressures, indicated for each curve. *b*, Resistance curves for sample C as a function of inverse temperature for different pressures indicated for each curve. The curves for 18.8 GPa and 21 GPa have been shifted to avoid crossings, as their real magnitudes coincide with those of the curves at 15.5 GPa and 11.2 GPa, respectively. Insets: logarithmic derivative of the resistance with respect to inverse temperature to show the temperature variation of the gap for both samples.

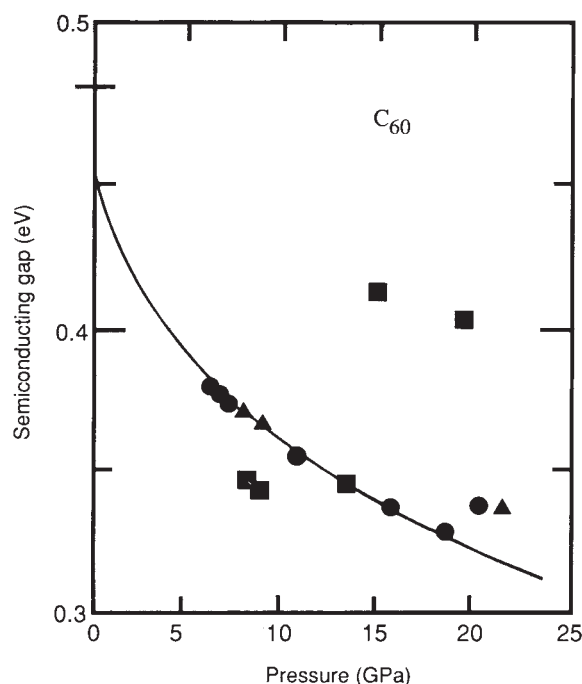


FIG. 2 Pressure dependence of the gap as extracted from the resistance: $\blacktriangle$, sample A; $\blacksquare$, sample B; $\bullet$, sample C. The value at ambient pressure (+) is taken from ref. 8; and the solid line represents the scaling from the equation of state (ref. 5).

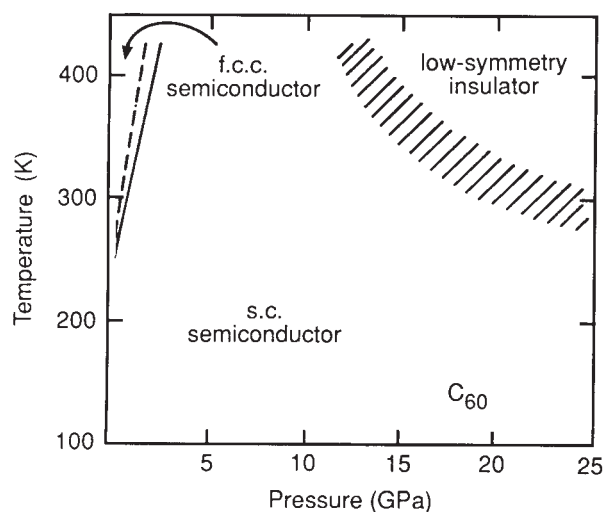


FIG. 3 Tentative phase diagram for solid C_{60} . Dashed line, data from ref. 11; solid line, data from ref. 12.

indicate that a semiconductor-metal transition occurs near 0.2 TPa. We find, however, that it is more favourable energetically to the system to become insulating above 20 GPa. This behaviour can be tentatively attributed to the formation of a covalent solid with a high degree of sp^3 bonding between different C_{60} clusters, similar to that found in diamond. $\square$

5. Duclos, S. J., Brister, K., Haddon, R. C., Kortan, A. R. & Thiel, F. A. *Nature* **351**, 380–382 (1991).
6. Kraetschmer, W., Lamb, L. D., Fostiropoulos, L. D. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
7. Haufier, R. E. *et al.* *J. Phys. Chem.* **95**, 8634 (1990).
8. Mort, J., Ziolo, R., Machonkin, M., Huffman, D. R. & Ferguson, M. I. *Chem. Phys. Lett.* (in the press).
9. Heiney, P. A. *et al.* *Phys. Rev. Lett.* **66**, 2911–2914 (1991).
10. Tolbert, S. H., Alivisatos, A. P., Lorenzana, H. E., Kruger, M. B. & Jeanloz, R. *Chem. Phys. Lett.* (submitted).
11. Kriza, G. *et al.* *J. Phys. I (France)* **1**, 1361 (1991).
12. Samara, G. A. *et al.* *Phys. Rev. Lett.* (in the press).
13. Appel, J. & Lautz, G. *Physica* **20**, 1110 (1954).
14. Bube, R. H. *Phys. Rev.* **106**, 703 (1957).
15. Gressier, P., Meerschaut, A., Guemas, L., Rouxel, J. & Monceau, P. *J. Solid St. Chem.* **51**, 141–151 (1984).

Electrically erodible polymer gel for controlled release of drugs

Ick Chan Kwon, You Han Bae & Sung Wan Kim

Center for Controlled Chemical Delivery, University of Utah,
421 Wakara Way, Salt Lake City, Utah 84108, USA

NEW controlled drug-delivery systems are being explored to overcome the disadvantages of conventional dosage forms¹. For example, stimulated drug-delivery has been used to overcome the tolerance problems that occur with a constant delivery rate, to mimic the physiological pattern of hormonal concentration and to supply drugs on demand^{1,2}. Stimuli-sensitive polymers, which are potentially useful for pulsed drug delivery, experience changes in either their structure or their chemical properties in response to changes in environmental conditions². Environmental stimuli include temperature^{3,4}, pH^{5,6}, light (ultraviolet⁷ or visible⁸), electric field^{9–12} or certain chemicals¹³. Volume changes of stimuli-sensitive gel networks are particularly responsive to external stimuli, but swelling is slow to occur^{14,15}. As well as being useful in the controlled release of drugs, such systems also provide insight into intermolecular interactions¹⁶. Here we report on a novel polymeric system, which rapidly changes from a solid state to solution in response to small electric currents, by disintegration of the solid polymer complex into two water-soluble polymers. We show that the modulated release of insulin, and by extension other macromolecules, can be achieved with this polymeric system.

Poly(ethyloxazoline) (PEOx) and either poly(methacrylic acid) (PMAA)¹⁷ or poly(acrylic acid)¹⁸ form complexes by means of intermolecular hydrogen bonding between carboxylic and oxazoline groups. There are two possible sites in an ethyloxazoline repeat unit for hydrogen bonding with carboxylic groups: the carbonyl oxygen and the nitrogen (Fig. 1). With a 1:1 ratio of repeat units, carbonyl oxygens are probably the dominant interactive group.

To prepare such a complex, 0.099 g PEOx (average M_r = 500,000) and 0.086 g PMAA (average M_r = 60,000) were each dissolved in 10 ml distilled water. These polymer solutions were mixed to give a 1:1 ratio of repeat units. The complex formed immediately, resulting in precipitation in a powder form below pH 5. The complex dissolved instantly above pH 5.4, but was incomplete after a month at pH 5.0. This was attributed to the deionization or ionization of the carboxylic groups of PMAA at different pH. This apparent discrepancy between the precipitation and dissolution pH values probably resulted either from a change in pKa for the PMAA carboxylic groups before and after complex formation, or from a difference in kinetics.

For swelling studies, the polymer complex was moulded into a disk. To accomplish this, the precipitate particles were filtered and then soaked for 1 hour in acetone/water (65/35 by volume) resulting in sticky swollen particles. The swollen polymer was integrated and compressed between two Teflon blocks to make a disk-shaped matrix of 2-mm thickness and 15-mm diameter and then dried in vacuo for three days. Before the study, the disks were pre-swollen in 0.9% saline solution. The saline solution was equilibrated with the atmosphere over 10 days and

Received 24 September; accepted 28 October 1991.

1. Saito, S. & Oshiyama, A. *Phys. Rev. Lett.* **66**, 2637–2640 (1991).
2. Fleming, R. M. *et al.* *Nature* **352**, 787–788 (1991).
3. Chang, K. J. *et al.* *Phys. Rev. Lett.* **54**, 2375–2378 (1985).
4. Ruoff, R. S. & Ruoff, A. L. *Nature* **350**, 663–664 (1991).

the resulting pH of the solution was 5.1 ± 0.15 . Equilibrium swelling was reached within three days and the equilibrium water content was 30% by weight.

The swollen matrix was attached to a woven platinum wire cathode which was 1 cm away from the anode and immersed

in 0.9% saline solution. The saline solution was constantly stirred with a magnetic stirrer and was exchanged for fresh solution every 15 min. When an electric current was applied, the solid matrix surface facing the cathode started to dissolve, and no complex reforming was found. Because the cathode produces hydroxyl ions by electrolysis of water, the local pH near the cathode increases, and hydrogen bonding between the two polymers is disrupted, causing disintegration of the polymer complex into two water-soluble polymers. Figure 2a shows the weight loss of the polymer matrix under a continuous 10-mA electric current. The rate of weight loss of the polymer matrix was constant until 80% of the initial weight had been lost. The linearity of weight loss with time implies that this process occurs through erosion of the surface facing the cathode. Zero-order kinetics are characteristic of surface erosion in slab-type polymer matrices¹⁹. The linearity was maintained until the matrix lost its shape. Figure 2b demonstrates surface erosion by applying an on-off step function of electric current. During the 'on step' the current was 10 mA. A stepwise weight loss was observed until 80% total weight loss was reached. These results indicate that the surface erosion of this polymer system can be controlled either in a stepwise or continuous fashion under electric stimulus.

To apply this polymer system to modulated solute release, we selected insulin as a model compound. Regular zinc insulin (10 mg) was suspended in 10 ml of the mixed polymer solution (pH 5.5), and the complex was formed by decreasing the pH of the suspension below pH 5 with 0.1 M HCl solution. The insulin-containing polymeric complex was then compressed into a disk as previously described. The loading content of insulin was $0.5 \pm 0.2\%$ ($n=3$) by weight. The disks were soaked in 0.9% saline solution for three days before applying the electric current. The amount of insulin released during the three-day equilibrium period was less than 4% of initial insulin load. Release studies were performed under 5 mA electric current, and the amount of insulin released was determined by radioimmuno-assay.

When we applied a step function of electric current to the insulin-loaded matrix, insulin was released in a stepwise manner until 70% had been released (Fig. 3). The large deviation in the release rate was attributed to defects causing irregular erosion of the insulin-loaded device. The release of dispersed insulin in the matrix was hardly detectable in the absence of electric current.

Although this work is still a long way from practical application, this concept could be used to develop implantable

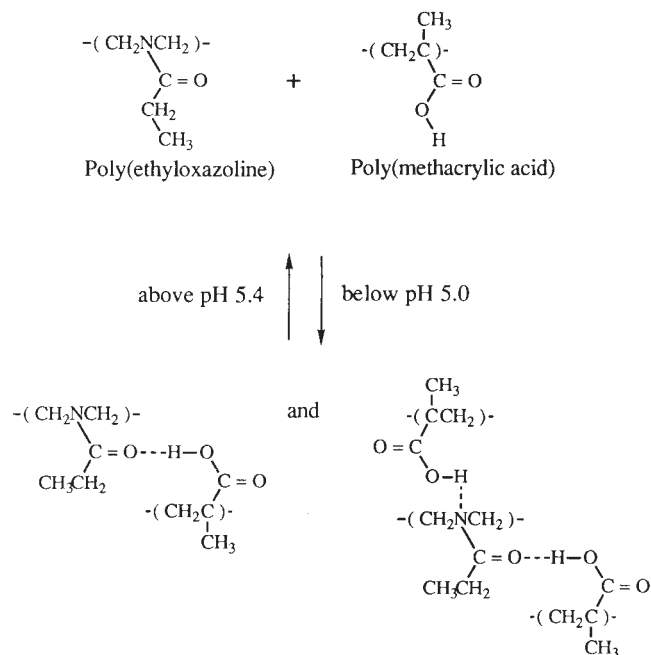


FIG. 1 Schematic illustration of the intermacromolecular interaction between poly(ethyloxazoline) and poly(methacrylic acid).

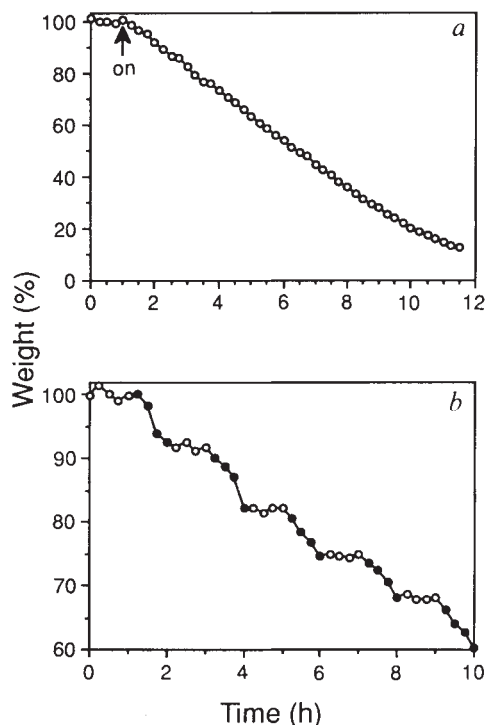


FIG. 2 Surface erosion of matrix consisting of poly(ethyloxazoline)-poly(methacrylic acid) complex in 0.9% saline solution when 10 mA electric current is passed. a, Continuous current. b, Step-function electric current (●, current on (10 mA); ○, current off).

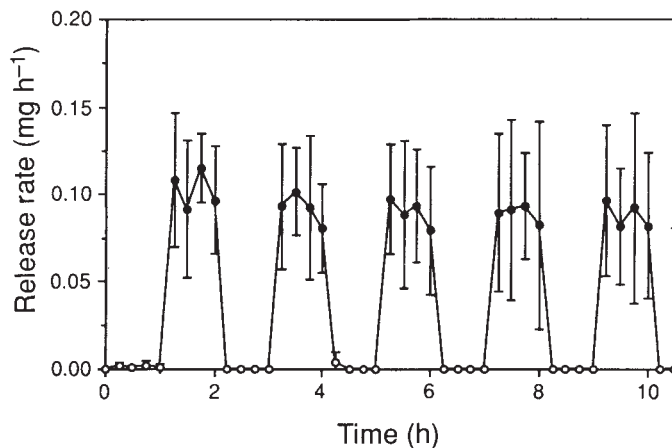


FIG. 3 Insulin release rate (normalized to mg h⁻¹ per 160 mg device) from insulin-loaded matrix of poly(ethyloxazoline)-poly(methacrylic acid) complex with the application of step-function electric current in 0.9% saline solution (mean $\pm \sigma$ from three measurements). ●, current on (5 mA); ○, current off.

therapeutic systems by considering polymer toxicity and structure, or used for isolated systems removed from physiological conditions, such as iontophoresis or a biosensor-linked reservoir. The approach could be used for potent polypeptide molecules, for which it is currently difficult to provide pulsed release. □

Received 17 July; accepted 11 October 1991.

1. Langer, R. *Science* **249**, 1527–1533 (1990).
2. Kost, J. (ed.) *Pulsed and Self-Regulated Drug Delivery* (CRC, Boca Raton, 1990).
3. Hoffman, A. S., Afrassabi, A. & Dong, L. C. *J. control. Release* **4**, 213–222 (1986).
4. Bae, Y. H., Okano, T., Hsu, R. & Kim, S. W. *Makromol. Chem. Rapid Commun.* **8**, 481–485 (1987).
5. Kuhn, W., Hargitay, B., Katchalsky, A. & Eisenberg, H. *Nature* **165**, 514–516 (1950).
6. Siegel, R. A. & Firestone, B. A. *Macromolecules* **21**, 3254–3259 (1988).
7. Ishihara, K., Hamad, N., Kato, S. & Shinohara, I. *J. Polym. Sci. Polym. Chem.* **22**, 881–884 (1984).
8. Suzuki, A. & Tanaka, T. *Nature* **346**, 345–347 (1990).
9. Tanaka, T., Nishio, I., Sun, S. & Ueno, S. *Science* **218**, 467–469 (1982).
10. Osada, Y. & Takeuchi, Y. *J. Polym. Sci. Polym. Lett.* **19**, 303–308 (1981).
11. Eisenberg, S. R. & Grodzinsky, A. J. *J. Membrane Sci.* **19**, 173–194 (1984).
12. Kwon, I. C., Bae, Y. H., Okano, T., Berner, B. & Kim, S. W. *Makromol. Chem., Macromol. Symp.* **33**, 265–277 (1990).
13. Ishihara, K., Muramoto, N. & Shinohara, I. *J. appl. Polym. Sci.* **29**, 211–217 (1984).
14. Hirokawa, Y. & Tanaka, T. *J. chem. Phys.* **81**, 6379–6380 (1984).
15. Ilmain, F., Tanaka, T. & Kokufuta, E. *Nature* **349**, 400–401 (1991).
16. Tsuchida, E. & Abe, K. *Adv. Polym. Sci.* **45**, 1–119 (1982).
17. Lichkus, A. M., Painter, P. C. & Coleman, M. M. *Macromolecules* **21**, 2636–2641 (1988).
18. Chen, F., Pearce, E. M. & Kwei, T. K. *Polymer* **29**, 2285–2289 (1988).
19. Hopfenberg, H. B. in *Controlled Release Polymeric Formulations* (eds Paul, D. R. & Harris, F. W.) Ch. 3 (American Chemical Society, Washington, DC, 1976).

ACKNOWLEDGEMENTS. We thank T. Okano and E. J. Mack for discussions.

Methane flux to the atmosphere from the Arabian Sea

N. J. P. Owens, C. S. Law, R. F. C. Mantoura,
P. H. Burkill & C. A. Llewellyn

Plymouth Marine Laboratory, Prospect Place, The Hoe,
Plymouth PL1 3DH, UK

ATMOSPHERIC concentrations of methane, an important greenhouse gas, have increased significantly over the past few decades^{1,2}. Although attention has been focused on anthropogenic sources, data from ice cores show that large changes in atmospheric methane concentrations have occurred over glacial–interglacial time scales, indicating that there is significant variability in natural methane fluxes^{3,4}. The surface waters of the oceans are often supersaturated with methane, which implies that the oceans are a net source, although their contribution to the global methane budget is small relative to other sources⁴. Here we report high concentrations of methane in the Arabian Sea, and calculate that the flux of methane to the atmosphere is up to five times greater than the previously reported average ocean flux. Methane production is associated with high phytoplankton biomass, which is closely coupled with the monsoon-driven upwelling of nutrient-rich water. We calculate that the Arabian Sea (representing 0.43% of the total surface area of the world's oceans) could account for between 1.3 and 133% of the current estimates of the open-ocean source of methane. Our results do not alter the view that the oceans are a relatively minor source of atmospheric methane, but the magnitude of the methane fluxes from the Arabian Sea and the link with the monsoon suggest that this region may be particularly sensitive to climate change, with a greater potential for feedback responses than its surface area might suggest.

A transect was followed from close to the Equator to the Gulf of Oman during September and October 1986, and dissolved methane was measured in the atmosphere, and from the sea surface to the bottom at 13 stations. Figure 1 shows that methane concentrations varied considerably with depth, being undersaturated, with respect to atmospheric concentrations, below 250 m at all stations. Such levels are typical of deep water^{5,6} and are thought to be largely the result of use by methane-oxidizing bacteria. Methane concentrations close to equilibrium

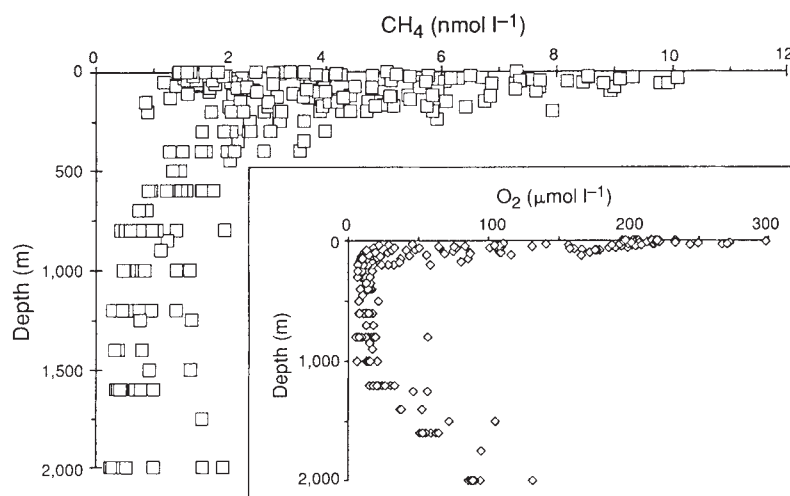
were found in the oligotrophic region to the south of the transect, and at stations 14 and 15, which had evidently (low temperature and high nitrate) been subject to upwelling of water of low methane concentration. However, much of the water column was supersaturated with methane at depths <200 m (Fig. 2). The supersaturations of methane extended to the surface, where the average saturation was 192% (s.d. = 34%; range 157–286%). These correspond to ΔCH_4 values (Fig. 2 legend) of 1.7 nmol l^{-1} (s.d. = 0.6, range 1.1–3.3 nmol l^{-1}). Supersaturated methane concentrations have been observed previously at a variety of locations^{6–10}, but the levels were considerably lower for non-sulphide-containing, offshore waters than those found during this study.

A strong pycnocline, with associated chlorophyll maxima, was apparent at 20–50 m depth at some stations (Fig. 3). Methane concentrations increased markedly at the pycnocline, and coincided with steep gradients in oxygen concentrations. The concentration of methane did not continue to increase with decreasing oxygen concentrations, which is normally observed in anoxic water masses, indicating that these were not the source of methane (compare with the Caribbean and the Black Sea)⁷. Instead, our profiles are similar to those of other open ocean sites where a peak in methane concentration has been observed at shallow depths^{8,9}.

The presence of supersaturated methane concentrations, in the essentially aerobic environment of the upper ocean, is enigmatic because methanogenesis is a bacterially mediated process, conventionally regarded as requiring strict anaerobic conditions. Various mechanisms for the formation of methane in the upper ocean have been suggested; for example, production in the guts of zooplankton¹¹, and within the low-oxygen interior of particles¹². Previous studies have also speculated about the possible links between phytoplankton and methane^{8,9,13}. In our study, methane exhibited an asymptotic relationship (probability level 0.00019) with chlorophyll ($\text{CH}_4 = 0.61 \times \log_{10} \text{chlorophyll } a + 4.65$; $r = 0.356$, $n = 104$), suggesting a coupling of methanogenesis with phytoplankton. This link was strongly supported by the production of methane from material collected from sediment traps (Fig. 4). As the traps were located at 50 m, in water depths of 1–3 km, the material evidently originated from recent biological production, with no contribution from sediment resuspension. Although these experiments do not provide information on the mechanism of methanogenesis, or its rate *in situ*, the observation is important as it indicates, unequivocally, that the potential for methane production was present in the upper water column, and was associated with sedimenting phylogenetic material. We suggest that the high concentrations of methane, *in situ*, were due, at least in part, to the close proximity of oxygen-depleted waters to the chlorophyll maximum, which would have provided both a source of organic material and a sufficiently reduced environment for methanogenesis. Oxygen levels were not regulated or measured in the shipboard trap incubation experiments, but the levels of oxygen undoubtedly declined over the period of incubation from fully aerated conditions at the start of the experiment. This does not reduce the validity of the experiments because, although low oxygen was apparently a requirement for methanogenesis, the highly reduced conditions of the oxygen minimum zone did not lead to methane concentrations increasing with depth, which would have been the case had the observed methane been produced by 'conventional' methanogenesis. Thus these observations add to the evidence for an upper-ocean source of methane.

The possibility that the methane was advected into the area from the Persian Gulf is excluded because these inputs were identified below 200 m (see also ref. 14). Also, the methane maximum was at station 9, and not 11, which would have been the case if the source was to the north. The possibility of lateral transport of methane produced in shallow-water sediments must also be considered. By assuming values for horizontal eddy

FIG. 1 Vertical distribution of dissolved methane and oxygen in the Arabian Sea. The data from all stations (Fig. 2) have been used, but only depths shallower than 2,000 m are shown. Samples were collected using a Neil Brown conductivity-temperature-depth system fitted with 12 × 10 l 'Go-Flo' sampling bottles. Samples were collected on the upcast and sub-sampled for gases immediately under N₂. Methane was extracted immediately from sub-samples (500 ml) by a static vacuum technique²² and 1 ml aliquots of the concentrated 'headspace' analysed for methane using a Varian gas chromatograph fitted with a flame ionisation detector. (OD Poropak Q column 2 m × 3 mm), ambient injector and column oven, detector 320 °C. Carrier, high-purity N₂ 30 ml min⁻¹ flow rate.) The analyses were calibrated using certified gas mixtures, and the standard error ((σ /mean) × 100) of the complete extraction and gas chromatogram procedure determined from repeated measurements of the same sample was 2.5% at 4 nmol l⁻¹; the detection limit (twice the background) averaged 0.3 nmol l⁻¹. Dissolved oxygen was measured using an automated Winkler titration technique²³, precision <1% at levels to 50 μ mol l⁻¹, detection limit 5 μ mol l⁻¹.



diffusion of $1 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$ (ref. 15) for a horizontal scale of 100 km, a source of methane in coastal waters of $1 \mu\text{mol l}^{-1}$ and a concentration of zero at our sites (a maximum gradient), and a flux to the atmosphere (see below) of $1 \text{ nmol cm}^{-2} \text{ day}^{-1}$, we calculate that horizontal diffusion could only account for ~5% of the observed concentrations of methane at station 11. This is a maximum because of the generous values for the variables chosen, and because biological removal (methane oxidation) has been ignored. Horizontal advection is impossible to estimate since the circulation of the area is poorly known and complex¹⁶. However, the direction of the surface currents and mass trans-

port in the top 300 m during the period of our study is to the south^{14,17} (parallel to our main transect). We thus conclude that, although there may have been a contribution to the methane maxima by transport of shelf-derived methane, the combined evidence of the coincidence of the methane maxima with (1) the oxycline, (2) the overlying chlorophyll maxima and (3) the methanogenic potential of sedimenting organic material, strongly supports an *in situ* source for the observed methane concentrations.

Although the presence of high methane concentrations in near-surface oceanic waters is of considerable intrinsic interest,

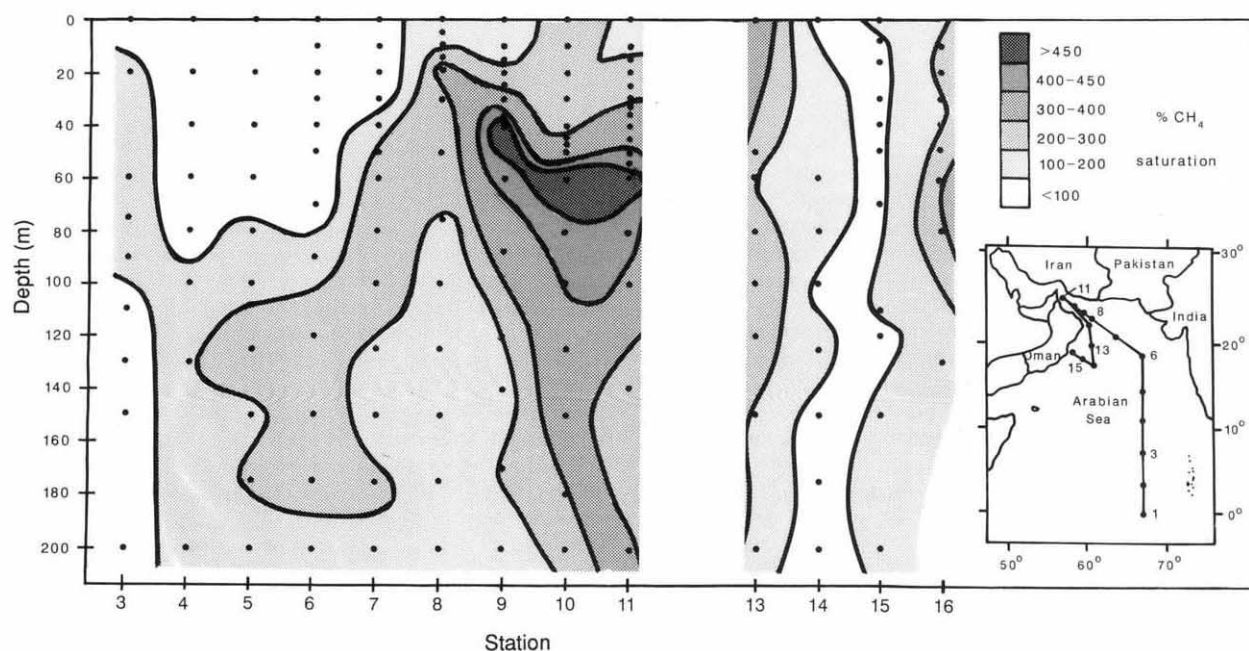


FIG. 2 Depth distribution of percentage saturation of dissolved methane in the Arabian Sea. Inset shows the station locations. Dissolved methane was measured as described in Fig. 1. The percentage saturation was obtained by measuring the atmospheric methane concentration (mean 77 pmol ml^{-1}) and calculating the expected concentration at the ambient salinity and temperature from the Bunsen solubility coefficients of Wiesenburg & Guinasso²⁴ and comparing with the observed concentration. Atmospheric fluxes were calculated by the expression: $\text{flux} = \Delta\text{CH}_4 \times \text{TV}$, where ΔCH_4 , in nmol l^{-1} , is the difference between the observed and expected concentration of methane and TV is the transfer velocity. TV is based on a linear interpolation of the wind speed and transfer velocity relationship in ref. 24. The maximum concentration of methane was 10.1 nmol l^{-1} (485% saturated) at 35 m at station 9. The average saturation in the surface 50 m for the

whole section was 208% (s.d. = 120%), and for the upwelling region (stations 8–16, excluding 14 and 15) 282% (s.d. = 93%). This region shows two important features: (1) a strong upwelling off the Arabian peninsula driven by the southwesterly monsoon (June–September); and (2) a zone of persistent oxygen depletion between 300–1,000 m. Our study was done at the end of the monsoon period, and the effects of the upwelling were apparent with nitrate concentrations and phytoplankton production generally increasing from the oligotrophic region in the south of the transect to the upwelling zone in the north^{25,26}. Stations 1–6 were characterized by low inorganic nutrients (for example $\text{NO}_3 < 20 \text{ nmol l}^{-1}$) and primary production ($< 0.3 \text{ mg C m}^{-2} \text{ d}^{-1}$), whereas stations 7–16 were influenced to varying degrees by upwelling, with nitrate concentrations up to $10 \mu\text{mol l}^{-1}$ and primary production in excess of $1 \text{ g C m}^{-2} \text{ d}^{-1}$ at some sites.

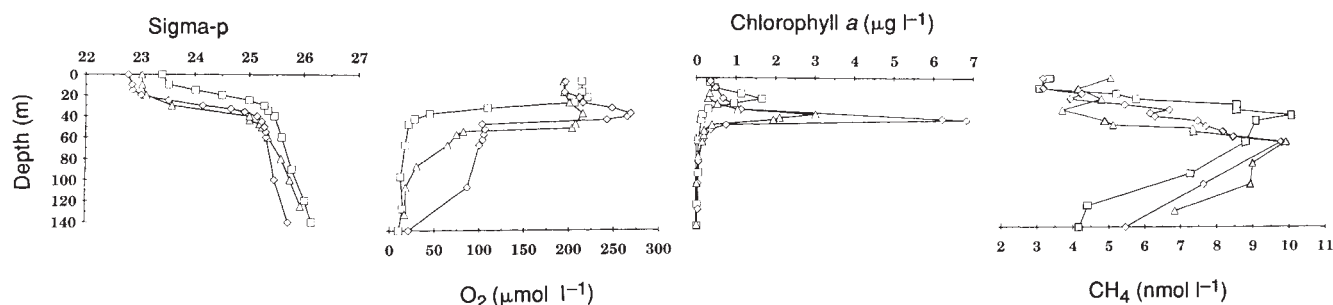


FIG. 3 Depth distribution of oceanographic variables at stations 9 (squares), 10 (diamonds) and 11 (triangles). Density (sigma-p) was obtained by conductivity-temperature-depth measurements, with salinity calibration by Guild-

line salinometer, and temperature by certified, digital thermometer. Chlorophyll *a* was measured by high-pressure liquid chromatography²⁷.

these observations are of greater significance because of the potential of the area as a source of methane. Assuming a minimum range of surface (0–10 m) ΔCH_4 for the upwelling region of $1.12\text{--}3.44\text{ nmol l}^{-1}$ (the average $\pm 1\sigma$), and a transfer velocity of 16.9 cm h^{-1} (Fig. 2 legend), we calculate fluxes of $0.46\text{--}1.39\text{ nmol cm}^{-2}\text{ day}^{-1}$, which are considerably greater (5–100 times) than those calculated for other oceanic regions^{6,9,10}. Transfer velocities show a nonlinear dependence on wind speed¹⁸. The transfer velocity used was based on an average wind speed of 10 m s^{-1} for the southwestern monsoon period (UK Meteorological Office). The wind speeds during this period

are frequently in excess of this (50% of the time $>12.5\text{ m s}^{-1}$), which would increase the transfer velocities considerably. In addition, the calculated fluxes would be considerably larger if the ΔCH_4 values averaged over the water column from the surface to the depth of maximum concentration were used instead of the 0–10 m interval. We consider it likely, therefore, that the actual fluxes from the Arabian Sea could be greater than those calculated.

Assuming that the average ΔCH_4 measured persists throughout the year, and a mean transfer velocity of 8.5 cm h^{-1} (for an annual average wind speed of 5.8 m s^{-1} ; UK Meteorological Office), we calculate an annual flux of methane to the atmosphere from this area of 0.04 Tg . This compares with the most recent estimates of the global oceanic (not including coastal sources) methane flux of $0.03\text{--}3\text{ Tg yr}^{-1}$ (ref. 10). Thus, the northern Arabian Sea, which represents 0.43% ($1.56 \times 10^6\text{ km}^2$ surface area) of the total surface area of the ocean, can contribute between 1.3 and 133% of the total oceanic flux of methane. Although this is small compared with the total global flux ($\sim 540\text{ Tg yr}^{-1}$ from all sources⁴), given the current debate about the global methane budget¹⁹ these data imply that the current oceanic flux should be revised upwards and point to the need to evaluate other potential oceanic sources.

It is evident that the northern Arabian Sea is an oceanic region of unusually high methane concentrations and fluxes to the atmosphere. Primary production in this region is amongst the highest of any ocean¹⁶. Sedimentation rates are also high and closely linked to primary production and the degree of monsoon-driven upwelling²⁰. The same region has also been shown to be a significant source of nitrous oxide²¹, and we suggest that if accurate values of the oceanic source of radiatively active gases are to be obtained, large-area averages are inappropriate; rather, the marked heterogeneity of the sources must be taken into account. Furthermore, as the source of methane appears to be, at least in part, phytogenic, the disproportionate importance of the area, and link with the monsoons, both suggest that the Arabian Sea may be particularly sensitive to climate change, with the potential for feedback responses greater than its surface area suggests. These data are also further evidence that biologically mediated ocean-atmosphere interactions occur through relatively discrete areas of activity. □

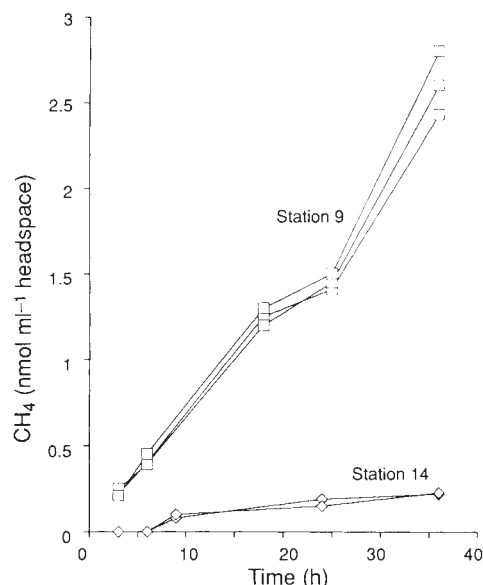


FIG. 4 Rate of methane production from material collected in sediment traps. Sediment traps (Institute für Meereskunde, Kiel) were deployed at 50 m for 24 h. After recovery, the trapped material was resuspended in filtered sea water and quantitatively distributed into several fractions (B. Zeitzschel, unpublished work). Aliquots 2 ml were placed in 5 ml serum bottles and sealed. The bottles were incubated in the dark at 25°C (close to ambient sea-water temperature). After gentle shaking, 0.25 ml of headspace was removed at intervals and analysed for methane. Atmospheric pressure was maintained by the addition of an equal volume of sea water. The vertical flux of material calculated from the traps was $46\text{ mg C m}^{-2}\text{ day}^{-1}$ at station 9, and $29\text{ mg C m}^{-2}\text{ day}^{-1}$ at station 14 (B. Zeitzschel, unpublished data). The aliquots incubated were (0.26 mg 0.06 mg C) per bottle for station 9 and 0.12 mg (0.04 mg C) per bottle for station 14. One aliquot of material from each station was incubated with the addition of gluteraldehyde (biological control), water only controls were also incubated. No methane production was detected in the controls. Similar experiments were done at station 5, and no methane production was observed.

Received 8 April; accepted 30 September 1991.

1. Rasmussen, R. A. & Khalil, M. A. K. *J. geophys. Res.* **86**, 9826–9832 (1981).
2. *Climate Change, The IPCC Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 18–22 (Cambridge University Press, Cambridge, 1990).
3. Chappelaz, J., Barnola, J. M., Raynaud, D., Korotkevich, Y. D. & Lorius, C. *Nature* **345**, 127–131 (1990).
4. Cicerone, R. J. & Oremland, R. S. *Global biogeochem. Cycles* **2**, 299–327 (1988).
5. Scranton, M. I. & Brewer, P. G. *Limnol. Oceanogr.* **23**, 1207–1213 (1978).
6. Ward, B. B., Kilpatrick, K. A., Novelli, P. C. & Scranton, M. I. *Nature* **327**, 226–229 (1987).
7. Atkinson, L. P. & Richards, F. A. *Deep-Sea Res.* **14**, 673–684 (1967).
8. Scranton, M. I. & Farrington, J. W. *J. geophys. Res.* **82**, 4947–4953 (1977).
9. Scranton, M. I. & Brewer, P. G. *Deep-Sea Res.* **24**, 127–138 (1977).
10. Conrad, R. & Seiler, W. *Deep-Sea Res.* **35**, 1903–1917 (1988).

11. Oremland, R. S. *Limnol. Oceanogr.* **24**, 1136–1141 (1979).
12. Sieburth, J. McN. *Prog. Oceanogr.* **21**, 117–128 (1988).
13. Lamontagne, R. A., Swinnerton, J. W., Linnenbom, V. J. & Smith, W. D. *J. geophys. Res.* **78**, 5317–5324 (1973).
14. Wyrtki, K. *Oceanographic Atlas of the International Indian Ocean Expedition*, 531 (National Science Foundation, Washington, DC 1971).
15. Okubo, A. *Deep-Sea Res.* **18**, 789–802 (1971).
16. Qasim, S. Z. *Deep-Sea Res.* **29**, 1041–1068 (1982).
17. Krey, J. & Babenerd, B. *Atlas of the International Indian Ocean Expedition. Phytoplankton Production* (IOC, UNESCO, Paris, 1976).
18. Watson, A. J., Upstill-Goddard, R. C. & Liss, P. S. *Nature* **349**, 145–147 (1991).
19. Crutzen, P. J. *Nature* **350**, 380–381 (1991).
20. Nair, R. R. et al. *Nature* **338**, 749–751 (1989).
21. Law, C. S. & Owens, N. J. P. *Nature* **346**, 826–828 (1990).
22. Scranton, M. I. & McShane, K. *Cont. Shelf Res.* **11**, 37–52 (1991).
23. Williams, P. J. LeB. & Jerkinson, N. W. *Limnol. Oceanogr.* **27**, 576–584 (1982).
24. Wiesenburg, D. A. & Guinasso, N. L. Jr. *Chem. Eng. Data* **24**, 356–360 (1979).
25. Owens, N. J. P. et al. *Deep-Sea Res.* (in the press).
26. Mantoura, R. F. C. et al. *Deep-Sea Res.* (in the press).
27. Mantoura, R. F. C. & Llewellyn, C. A. *Analyt. Chim. Acta* **151**, 297–314 (1983).

ACKNOWLEDGEMENTS. We thank the captain, officers and crew of the RRS *Charles Darwin* for service at sea, and the colleagues who assisted us during the cruise, particularly J. Morris for the oxygen measurements. We thank B. Zeitschel for supplying the sediment trap material and R. J. Uncles for advice on the physical oceanography of the area. N.J.P.O. is grateful to J. McN. Sieburth for initiating an interest in methane cycling and the Royal Society for a travel grant. The Plymouth Marine Laboratory is part of the UK NERC.

Rapid dewatering of the crust deduced from ages of mesothermal gold deposits

R. J. Goldfarb*, L. W. Snee*, L. D. Miller† & R. J. Newberry‡

* US Geological Survey, Box 25046, MS 973, Denver Federal Center, Denver, Colorado 80225, USA

† Echo Bay Mines, 3100 Channel Drive, Juneau, Alaska 99801, Canada

‡ Department of Geology, University of Alaska, Fairbanks, Alaska 99775, Canada

THE large-scale migration of fluids through the continental crust has been well documented, but there is no consensus regarding the timing of fluid migration relative to orogenic episodes, or rates of crustal dewatering¹. Here we present ⁴⁰Ar/³⁹Ar dates for muscovites from quartz veins along a major shear zone in southeast Alaska, which show that the veins were emplaced in the early Eocene, during the late stages of orogenic deformation. Hydrothermal activity took place for only about 1 Myr and along a distance of at least 200 km. The fluids were generated by metamorphic reactions in subducted crust along the North American plate margin, and were apparently trapped in the crust by the low permeabilities accompanying a convergent tectonic regime until 56 Myr ago. The rapid dewatering event coincided with a change in plate motion at 56–55 Myr, which caused a shift from convergent to partly transcurrent tectonics. We suggest that this change in tectonic regime led to increased crustal permeabilities and hence the possibility of large-scale fluid migration.

Terrane accretion and subduction, thrusting, deformation, regional metamorphism and magmatic thickening of the Coast orogen characterized the northern North American cordillera between 110 and 50 Myr (ref. 2). Rapid uplift of the orogen core occurred during the last 15 million years of this period³. The 700-km-long, northwest-trending Coast Range megaclineament cuts medium- and high-grade metamorphic rocks 15–20 km west of the core of the orogen and of the Coast batholith in southeast Alaska and British Columbia. The megaclineament has been defined in Alaska and British Columbia. The megaclineament has been defined in Alaska as a prominent topographic zone, up to 10 km wide, formed by selective erosion along closely spaced joints, foliation, compositional layering and small faults⁴. Farther south in British Columbia, this structure is recognized as a steep to vertical ductile shear zone⁵. There the megaclineament is interpreted to be the westernmost of several ductile

shear zones that formed between 65–55 Myr and 50–48 Myr as a result of rapid differential uplift of the Coast orogen².

Gold-bearing quartz veins of the Juneau gold belt are widespread along the northern 200 km of the megaclineament. These veins record a high fluid flux along this structure. Stable isotope data from silicate minerals in the veins can be used⁶ to calculate $\delta^{18}\text{O}$ values for the ore fluid of 7.2–12.8‰ and δD values of –35 to –15‰ $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}}] - 1$ and $\delta\text{D} = [(^2\text{D}/^1\text{H})_{\text{sample}} / (^2\text{D}/^1\text{H})_{\text{standard}}] - 1$. Fluid inclusion data⁷ indicate that the hydrothermal fluids were rich in H_2O , CO_2 , CH_4 , N_2 and H_2S , and that the veins were formed at depths of at least 5 km and at temperatures of at least 250 °C. These deposits are typical of mesothermal gold-bearing vein systems found throughout the Earth's orogenic belts.

Many different sources for the vein-forming fluids can be postulated, but existing data are only consistent with a metamorphic fluid. Interstitial pore fluids in the subducted, sedimentary rock-dominant terranes would have been expelled at shallow depths, long before the host rocks of the veins reached their maximum burial depths of perhaps 30–35 km (ref. 8). The lack of temporal and spatial association of the veins with specific igneous activity, and the consistently low salinity of the fluids⁷, rule out the possibility of a magmatic fluid source. Calculated δD values for the fluid are incompatible with a meteoric water source with any feasible water/rock regime⁶. The stable-isotope and fluid-volatile chemistry, as well as the fact that the vein swarms are restricted to greenschist facies rocks, suggest that the fluid was generated during metamorphic reactions of greenschist and amphibolite facies^{6,7}.

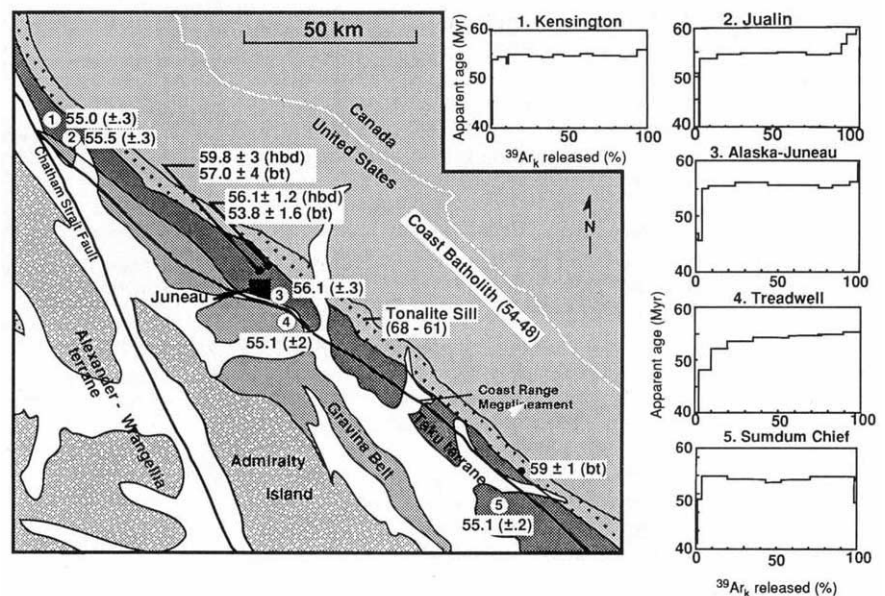
The ⁴⁰Ar/³⁹Ar dates of hydrothermal muscovite from the five largest vein swarms in the gold belt show a remarkably narrow range, 55.0–56.1 Myr, for samples collected along 200 km (Fig. 1). Well-defined plateaus characterize all spectra of ³⁹Ar release. The concordance of apparent ages might indicate that the dates of micas had been reset by a single thermotectonic episode or cooling of all veins through the 325 °C isotherm at ~56–55 Myr. It could, however, indicate actual fluid ascension and vein deposition at geologically rapid rates; that is, extensive metamorphic dewatering and gold vein formation may be a relatively brief and late event during a much more prolonged period of orogenesis.

The resetting of all ⁴⁰Ar/³⁹Ar dates to 56–55 Myr from originally older dates is unlikely. Exposed igneous rocks in the immediate vicinity of the Juneau gold belt are clearly older than quartz veins. Veins cut through 105-Myr monzodiorite⁹ at Kensington and Jualin, Cretaceous monzodiorite at Treadwell and Mesozoic metagabbro at Alaska–Juneau. A belt of tonalitic plutons, generally 5 km east of the megaclineament, was episodically emplaced from 68.2 ± 2.0 to 61 ± 1.5 Myr (refs 10–12). Granite and granodiorite of the Coast batholith were intruded between 54 ± 2 and 48 ± 2 Myr within a belt 20–50 km east of, and parallel with, the megaclineament¹³. This belt of plutons could in theory have reset the hydrothermal micas to 56–55 Myr, but the thermal effects of the magmatism apparently did not extend as far west as the megaclineament (fig. 1). Hornblende and biotite K/Ar data for schist and amphibolite ~5 km east of the structure at the latitude of Juneau show no evidence of thermal overprinting¹⁴. Farther south, near the Sumdum Chief mine, Ar–Ar studies¹² indicate that the western margin of the belt of tonalite bodies has cooled below the 280 ± 40 °C biotite blocking temperature by 59 ± 1 Myr.

It is unlikely that the micas were reset by circulation of meteoric water during uplift. Silicate ΔD values for country rocks near the megaclineament attest to the lack of significant meteoric water circulation west of the Coast batholith⁶. Furthermore, even if such circulation did occur, the undisturbed shape of the ³⁹Ar spectra (Fig. 1) is evidence against resetting.

It is also doubtful that the 56–55 Myr concordance reflects uplift of previously formed veins across the 325 ± 25 °C isotherm, which is the approximate blocking temperature for sericite¹⁵.

FIG. 1 Dates from $^{40}\text{Ar}/^{39}\text{Ar}$ data for hydrothermal muscovite from gold-bearing veins at the five largest mines of the Juneau gold belt, ranging from 56.1 to 55.0 Myr. Ages obtained from the U/Pb ratio are 68–61 Myr (refs 8–10) for a tonalite sill belt to the east of the gold belt, indicating that it pre-dates the gold veining. U/Pb ages for the Coast batholith¹¹ provide evidence that some magmatism may have been coeval with ore deposition. But K/Ar dates of metamorphic rocks to the east of Juneau¹⁴ and $^{40}\text{Ar}/^{39}\text{Ar}$ dates from the southern part of the sill belt¹² suggest that (1) the thermal effects of Eocene igneous activity were not felt as far west as the gold belt and (2) rocks near the gold belt had already cooled below the biotite blocking temperature of 280 °C before gold was formed. Dates labelled 'hbd' are from hornblende; bt, biotite.



Fluid inclusion studies⁷ suggest that vein formation temperatures were probably less than or close to 325 °C. Lower greenschist facies rocks¹⁶ that host the veins at the Treadwell mine on the west side of the megalinearment never reached temperatures above ~325 °C. The uplift histories were probably notably different for rocks on opposite sides of the megalinearment^{2,17}, and there is no evidence that the sampled veins on both sides of the structure cooled through the 325 °C isotherm at exactly the same time.

The most logical explanation for the exceptional concordance of calculated dates is that massive fluid circulation and vein formation occurred over a very brief time span in the early Eocene. This event took place over a length of at least 200 km along the Coast Range megalinearment. Whether such an event affected other parts of the megalinearment is uncertain. The lack of extensive gold-bearing quartz vein networks along the southern part of the structure might be explained by the fact that deeper crustal exposures of the megalinearment and surrounding rocks occur to the south. If such veining did occur to the south, the evidence has long since been eroded away.

The widespread fluid event at 56–55 Myr might reflect an important change in crustal stress regime, as considerable changes in plate motion occurred at roughly that time (Fig. 2). Plate trajectory models¹⁸ show a counterclockwise swerve of the Kula plate concentrated in the 56–55 Myr interval, causing a shift from orthogonal subduction to a more oblique direction and the initiation of strike-slip motion. Structural and uplift studies¹⁹ along the southern part of the megalinearment indicate that compressional tectonic events ceased between 60 Myr and 55 Myr. An abrupt change in the type and location of magmatism occurred near the megalinearment between 54 and 48 Myr. The locus of magmatism shifted 20–50 km farther to the east-northeast, and plutons became more siliceous in composition and characterized by lower initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, pointing to changes in magma source¹³. It was suggested¹³ that these characteristics correlated with the change to a more extensional tectonic regime. The emplacement of the siliceous plutons in response to changing crustal stresses may be a more gradual process than fluid migration and quartz vein emplacement.

Our favoured scheme (Fig. 2) is one in which the vein-forming fluids may have formed over millions of years during devolatilization of subducted crust. A period of progressive Barrovian metamorphism from 70 Myr to ~60–55 Myr, superimposed on deeply buried, high-pressure, low-temperature prehnite-pumpellyite facies and lower greenschist facies rocks east of the megalinearment⁸, is the event likely to have produced the large

fluid volumes. The cause of this metamorphic event is uncertain. Re-equilibration of perturbed isotherms during uplift and/or during a declining rate of convergence^{20,21} is one possibility. Alternatively, thermal input from intrusion of the tonalites has been suggested⁸.

Rather than migrating upwards along the megalinearment at various times, fluids might have ponded at depth below impermeable units. The mainly compressional environment kept permeabilities low along the western edge of the Coast orogen, even along any existing principal shear zones, including the

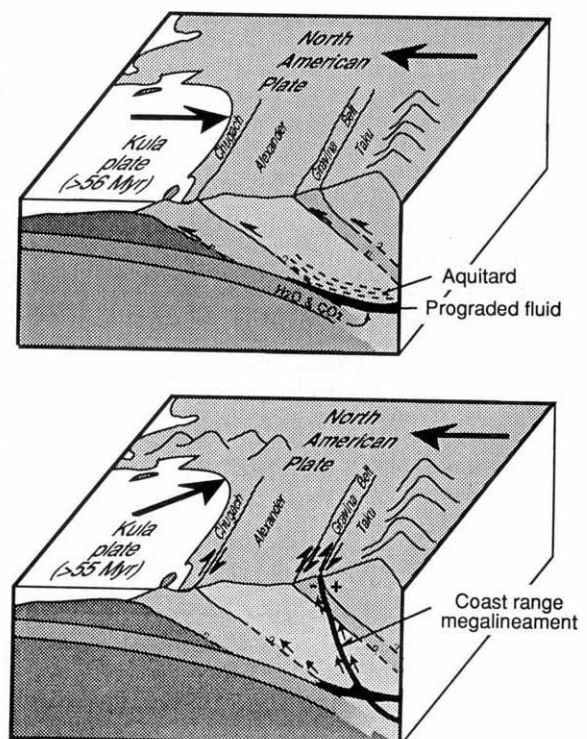


FIG. 2 Vein-forming fluids were released through devolatilization of subducted sedimentary and volcanic rocks during plate convergence before 56–55 Myr. Compressional tectonics maintained low crustal permeabilities, and therefore the released fluids probably collected below relatively impermeable units. A shift to a more oblique plate margin regime at 56–55 Myr (ref. 18) initiated considerable strike-slip motion and allowed hydrothermal fluids to ascend rapidly.

Coast Range megalineament (if it then existed). Seismic reflection data from present-day convergent margins provide evidence that large fluid volume may be trapped at mid-crustal levels during compressional events²². Hydrologic modelling²³ of metamorphic fluids supports the seismic observations and indicates significant crustal porosities even long after peak metamorphism. The hydrology of deep crustal fluids is complex, however, and needs further study.

The trapped fluids were rapidly discharged along the megalineament over a period of $\leq 10^6$ yr. Failure along the structure during some degree of relaxation of compression at 56–55 Myr reduced confining pressures and created pockets of fluid overpressure, especially in zones of maximum dilation. Swarms of brittle–ductile shear veins²⁴ formed in dilational sites within a few kilometres of the megalineament during large-scale crustal failure, probably according to the fault valve model of Sibson and others²⁵.

Mesothermal gold-bearing quartz vein networks may delineate zones of metamorphic dewatering²⁶ and are often spatially associated with principal crustal lineaments. Dating of gold-bearing, metamorphic quartz veins located along large-scale structures allow precise determination of the time of changes in crustal stresses and possibly in relative plate movement. Dewatering of the middle crust is a relatively instantaneous consequence of seismic failure, perhaps marking the start of a more transpressive type of plate margin regime. The discharge of non-metamorphic fluids immediately following seismic events has been documented along many shear zones²⁷. The isotope dates of 56–55 Myr from the Juneau gold belt are believed to define the onset of failure along the northern part of the Coast Range megalineament.

The formation of other Phanerozoic gold systems in western North America may also correlate with periods of considerable shift in plate motion. A marked change in the relative motion of the Farallon and North American plates at 133 Myr may be recorded by gold veining in the Klamath mountains²⁸. Along the Melones fault zone of central California, K/Ar and Rb/Sr data from some of the mines in the Alleghany district, near the north end of the Melones fault zone, and from the Mother Lode district, ~200 km farther south on what may be the same structure, are statistically identical at 110–116 Myr (ref. 25). Plate reconstruction models³⁰ for the southern Cordillera often include a shift from left-oblique to right-oblique convergence beginning at ~115 Myr. Although there is little evidence for Cretaceous strike-slip motion along the southern Melones fault zone³¹, the change in plate slip direction still may have been the driving force for localized opening of dilational jogs along the fault zone.

Our geochronological results from southeast Alaska clearly show that an episode of considerable fluid flow occurred relatively rapidly over great crustal lengths. We believe that the remarkable correlation between the time of this episode and that of significant shifts in plate motions is not coincidental. Rather, catastrophic crustal fluid flow and gold genesis along the outboard flank of an orogenic belt seem to be inherent tectonic processes, ultimately related to changing plate stresses in Cordilleran-type environments. □

15. Snee, L. W., Sutter, J. F. & Kelly, W. C. *Econ. Geol.* **83**, 335–354 (1988).
16. Dusel-Bacon, C., Brew, D. A. & Douglass, S. L. *US Geol. Surv. Open-file Rep.* **91-96**, 46P (1991).
17. Stowell, H. H. & Hooper, R. J. *Tectonics* **9**, 391–407 (1990).
18. Lonsdale, P. *Geol. Soc. Am. Bull.* **100**, 733–754 (1988).
19. Crawford, M. L. & Crawford, W. A. *Geol. Assoc. Can. Progr. Abstr.* **15**, A28 (1990).
20. England, P. C. & Thompson, B. *J. Petrol.* **25**, 894–955 (1984).
21. Kerrich, R. & Wyman, D. *Geology* **18**, 882–886 (1990).
22. Hyndman, R. D., Yorath, C. J., Clowes, R. M. & Davis, E. E. *Can. J. Earth Sci.* **27**, 313–329 (1990).
23. Thompson, A. B. & Connolly, J. A. D. *Tectonophysics* **182**, 47–55 (1990).
24. Miller, L. D. & Redman, E. C. *Geol. Soc. Austr. Abstr.* **23**, 374 (1988).
25. Sibson, R. H., Robert, F. & Poulsen, K. H. *Geology* **16**, 551–555 (1988).
26. Henley, R. W., Norris, R. J. & Paterson, C. J. *Mineral. Deposita* **11**, 180–196 (1976).
27. Barnes, I., Irwin, W. P. & White, D. E. *US Geol. Surv. Misc. Invest. Map* I-1528 (1984).
28. Elder, D. R. & Cashman, S. M. *Geol. Soc. Am. Abstr. Prog. Cord. Sect.* **23**, 21 (1991).
29. Bohlke, J. K. & Kistler, R. W. *Econ. Geol.* **81**, 296–322 (1986).
30. Engbreton, D. C., Cox, A. & Gordon, R. G. *Geol. Soc. Spec. Pap.* **206**, 45P (1985).
31. Schweickert, R. A., Bogen, N. L. & Merguerian, C. *Metamorphism and Crustal Evolution of the Western United States*, Vol. VII (ed. Ernst, W. G.) 789–822 (1989).

ACKNOWLEDGEMENTS. We thank Fred Barker, Bob Burruss, Lang Farmer and D. Lamar Leach for reviews, Rick Fredricksen and Richard Kirkham for mine access, and Barrett Cieutat and J. Carter Borden for laboratory assistance.

Tetrapod or near-tetrapod fossils from the Upper Devonian of Scotland

P. E. Ahlberg

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

SINCE 1932, the earliest known undisputed tetrapods have been of uppermost Famennian (late Upper Devonian) age^{1–3}. Although a probable tetrapod jaw has been described from the Lower Famennian⁴, and tetrapod tracks of supposedly Frasnian⁵ (and possibly earlier⁶) age are known, no fossil limb material older than the latest Famennian has been discovered. The 'panderichthyids' *Panderichthys*⁷ and *Elpistostege*⁸ from the Lower Frasnian (early Upper Devonian) are regarded by some^{8–10} as the closest known sister group of tetrapods, but *Panderichthys* has paired fins rather than limbs¹¹. Here, I describe a hitherto unrecognized tibia from the Upper Frasnian (middle Upper Devonian) site of Scat Craig, near Elgin, Scotland^{12–14}, collected during the nineteenth century, which extends the fossil record of the tetrapod-type hind limb by roughly seven million years¹⁵. Other isolated bones from the same locality also show tetrapod characteristics. The tibia, a humerus and some incomplete jaws are discussed below, but a complete description of the material is in preparation.

The tibia (University Museum, Oxford, specimen OUM D793) lacks the proximal articular surface and posterior margin, but like the other specimens it is otherwise well preserved and undistorted (Fig. 1a–e). Because the proximal part of the anterior surface is flared, as would be expected near the knee joint, virtually the whole length of the bone appears to be preserved. In general OUM D793 resembles the tibiae of *Ichthyostega*^{2,16} and *Acanthostega*. But if the length is correctly reconstructed it is proportionately shorter (Fig. 1f, g), which may be a primitive characteristic. The cnemial crest, developed as a sharp edge between the anterior and extensor surfaces (Fig. 1b, d), is less prominent than in *Ichthyostega*. As the tibia carries articular surfaces that appear to be for intermedium and tibiale, it was presumably associated with a fully formed tarsus; this is regarded as a tetrapod autapomorphy^{17–19}.

The humerus (British Geological Survey, Nottingham, specimen GSM 104536) shows a mixture of derived tetrapod features^{16,19–22}, primitive characters shared with osteolepiforms^{23–25} (the probable sister group of panderichthyids and tetrapods^{8–10}), and apparently unique characters (Fig. 2a–f, Table 1). It is more tetrapod-like than any known fish humerus^{23–27} (Fig. 2g–i), but lacks the characteristic early tetrapod 'L'-shape^{2,16,19–22}, as the region proximal to the

Received 19 June; accepted 23 September 1991.

1. Torgersen, T. *EOS* **72**, 18–19 (1991).
2. Crawford, M. L., Hollister, L. S. & Woodsworth, G. J. *Tectonics* **6**, 343–361 (1987).
3. Hollister, L. S. *Can. Mineral.* **20**, 319–332 (1982).
4. Brew, D. A. & Ford, A. B. *Can. J. Earth Sci.* **15**, 1763–1772 (1978).
5. Crawford, M. L. & Hollister, L. S. *J. geophys. Res.* **87**, 3849–3860 (1982).
6. Goldfarb, R. J., Newberry, R. J., Pickthorn, W. J. & Gent, C. A. *Econ. Geol.* **86**, 66–80 (1991).
7. Goldfarb, R. J., Leach, D. L., Roses, S. C. & Landis, G. P. *Econ. Geol. Monogr.* **6**, 363–375 (1989).
8. Himmelberg, G. R., Brew, D. A. & Brew, D. A. *J. metamorphic Geol.* **9**, 165–180 (1991).
9. Gehrels, G. E. & Miller, L. D. *Conf. Juneau–91, Abstr. Prof. Pap.*, 23–24 (1991).
10. Gehrels, G. E., Brew, D. A. & Saleeby, J. B. *US Geol. Surv. Circ.* **939**, 100–102 (1984).
11. Barker, F., Arth, J. G. & Stern, T. W. *Am. Mineral.* **71**, 632–643 (1986).
12. Wood, D. J., Stowell, H. H., Onstott, T. C. & Hollister, L. S. *Geol. Soc. Am. Bull.* **103**, 849–860 (1991).
13. Barker, F. & Arth, J. G. *Geol. Soc. Am. Mem.* **174**, 395–405 (1990).
14. Forbes, R. B. & Engels, J. C. *Geol. Soc. Am. Bull.* **81**, 579–584 (1970).

TABLE 1 Comparison of humeral character states

Humerus	<i>Eusthenopteron</i>	GSM 104536	Tetrapods	Humerus	<i>Eusthenopteron</i>	GSM 104536	Tetrapods
Entepicondyle	Thick, curved, rounded end, distinct from cylindrical body of humerus	Thin, flat, rounded end, continuous with flat body of humerus	Thin, flat, rectangular, continuous with flat body of humerus	Pectoral process	Weak	Strong	Strong
Transverse humeral ridge ^{18,23}	Present, strong	<i>Absent</i>	Present	Proximal region	Short	Short	Long
Preaxial margin	Rounded	Sharp	Sharp	Ectepicondyle	Broad, low	Narrow, tall	Narrow, tall
				Ectepicondylar foramen	Present	<i>Absent</i>	Present*
				Supinator process	Close to ectepicondyle	Close to ectepicondyle	On preaxial margin
				Size of caput	Large	<i>Small</i>	Large

The Scat Craig humerus GSM 104536 was compared with those of the osteolepiform *Eusthenopteron*²³ and the early and primitive tetrapods *Ichthyostega*^{2,21}, *Acanthostega*¹⁶, *Crassigyrinus*²¹, *Proterogyrinus*²⁰, *Eschschschmidt*²² and *Greererpeton*¹⁹. The tetrapod humeri agree with respect to all the characters compared in the table, except (*) that an ectepicondylar foramen is absent in *Proterogyrinus*, *Eschschschmidt* and *Greererpeton*. (I regard this character state as derived within the Tetrapoda). The humerus of *Eusthenopteron* compares well with those of other osteolepiforms^{23–25} and rhizodonts²⁶, although there are minor differences in the shapes and positions of the processes; this humeral morphology is assumed to be primitive relative to that of tetrapods. 'Body' refers to the central core or shaft of the humerus, running from the caput to the radial and ulnar facets. The 'proximal region' is that part of the body of the humerus which lies proximal to the entepicondyle and pectoral process. Bold typeface indicates shared derived character state; italics, unique character state.

entepicondyle and pectoral process is very short. This seems to be a primitive, fish-like character, as are the rounded postaxial margin of the entepicondyle and the position of the supinator process close to the ectepicondyle^{19–22}. The most surprising features of GSM 104536 are the absence of a transverse humeral ridge, which suggests that the elbow flexor musculature was poorly developed, and the small caput humeri; these characters appear to be unique, but whether they are really autapomorphies remains uncertain. Little can be said about the distal parts of the appendage, as the radial and ulnar facets are not preserved.

Roughly 20 tetrapod-like mandibular fragments, carrying polyplacodont teeth^{18,28}, are known from Scat Craig. Most come from jaws of similar size to Royal Museum of Scotland specimen RMS G 1967.17.1 (Fig. 3c). All these apparently belong to one species, which, with a mandibular length estimated at 350–400 mm, seems to have been considerably larger than *Panderichthys* and nearly twice the size of *Ichthyostega*. A small jaw (Natural History Museum, London, specimen BMNH P62809, Fig. 3a) differs somewhat from the others in the proportions of its coronoid dentition, but as it agrees very closely with the other mandibles in tooth histology and general structure it more probably represents a subadult than a separate species. The Scat Craig jaws closely resemble two partial mandibles from the Upper Frasnian of Latvia, described under the name *Obruchevichthys*⁷ (Fig. 3b).

The lower jaws of osteolepiforms^{7,29} and panderichthyids³⁰ have broad coronoids with large fangs separated by fossae usually flooded by meckelian bone (Fig. 3d). In tetrapods^{2,20–22,31–33} (Fig. 3e–g), fangs and fossae are absent and the meckelian bone is not exposed dorsal to the prearticular: this condition is presumably derived. The Scat Craig mandibles are of tetrapod structure, but coronoid fangs are present although relatively small. Anterior to the first coronoid, a large parasymphysial (or 'adsymphysial'^{7,29}) plate (Fig. 3a, c) carries a fang pair and a continuation of the coronoid tooth row. Other

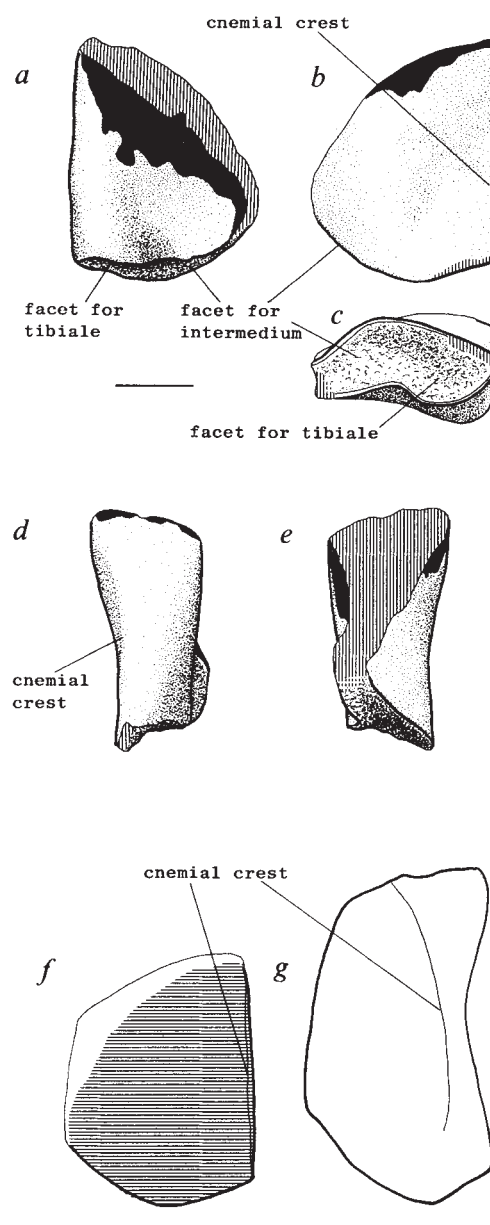


FIG. 1 a–e, Right tibia from Scat Craig, University Museum, Oxford, specimen OUM D793. a, Flexor (ventral) surface; b, extensor (dorsal) surface; c, distal surface; d, anterior (preaxial) surface; e, posterior (postaxial) view. Vertical hatching, broken bone; double cross-hatch, eroded but essentially true surface; thick line, true edge; thin line, broken edge. The distal surface consists of unfinished bone; the exact boundary between this surface and the broken posterior margin is uncertain. f, Reconstructed outline of OUM D793. Horizontal hatching, preserved bone. g, Tibia of *Ichthyostega* (after ref. 16, reversed). Scale bars, 10 mm.

than in the Scat Craig jaws, *Obruchevichthys* (Fig. 3b) and the probable tetrapod *Metaxygnathus*⁴ (P.E.A., personal observation) parasymphysial plates with comparable dentitions are only known in *Ichthyostega* (Fig. 3g) and an unnamed Carboniferous tetrapod³³ (Fig. 3e, f). In osteolepiforms²⁹ and *Panderichthys*³⁴ the plate carries shagreen and is usually smaller (Fig. 3d). The Scat Craig mandibles are in all respects more tetrapod-like than that of *Panderichthys*. Significantly, the former lack a ventral groove for the submandibular bones; the ossified gill cover may thus have been lost, as in tetrapods.

Three premaxillary fragments from Scat Craig, OUM D796, BMNH P9776 and RMS G 1891.92.444, can be associated with

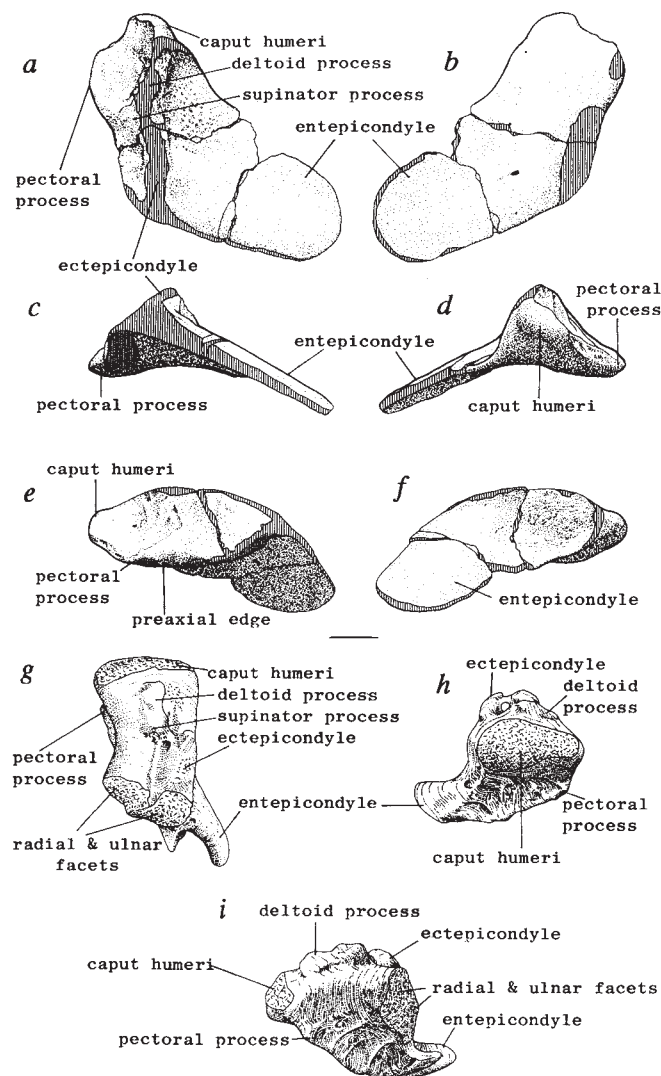


FIG. 2 a–f, Left humerus, British Geological Survey, Keyworth, Nottingham, specimen GSM 10456 (drawer 55/6). a, Dorsal; b, ventral; c, distal; d, proximal; e, preaxial; f, postaxial. Scale bar, 10 mm. Vertical hatching, broken bone; thick line, true edge; thin line, broken edge. The distal margin, ectepicondyle and entepicondyle have suffered considerable abrasion, and the radial and ulnar facets can no longer be recognized. Comparison with *Eusthenopteron* (University Museum of Zoology, Cambridge, specimen UMCZ GN190) suggests that the elongated area of broken bone along the preaxial margin of the ventral surface is the remains of a ridge which carried the radial facet at its distal end. Like many Scat Craig specimens, the bone was broken during excavation; the prominent notch in the proximal margin of the entepicondyle is due to breakage but may reflect (and have been caused by) the position of the entepicondylar foramen. By contrast, the absence of an entepicondylar foramen seems to be natural. g–i, Humerus of *Eusthenopteron* in dorsal (g), proximal (h) and preaxial view (i). From ref. 23. Scale bars, 10 mm.

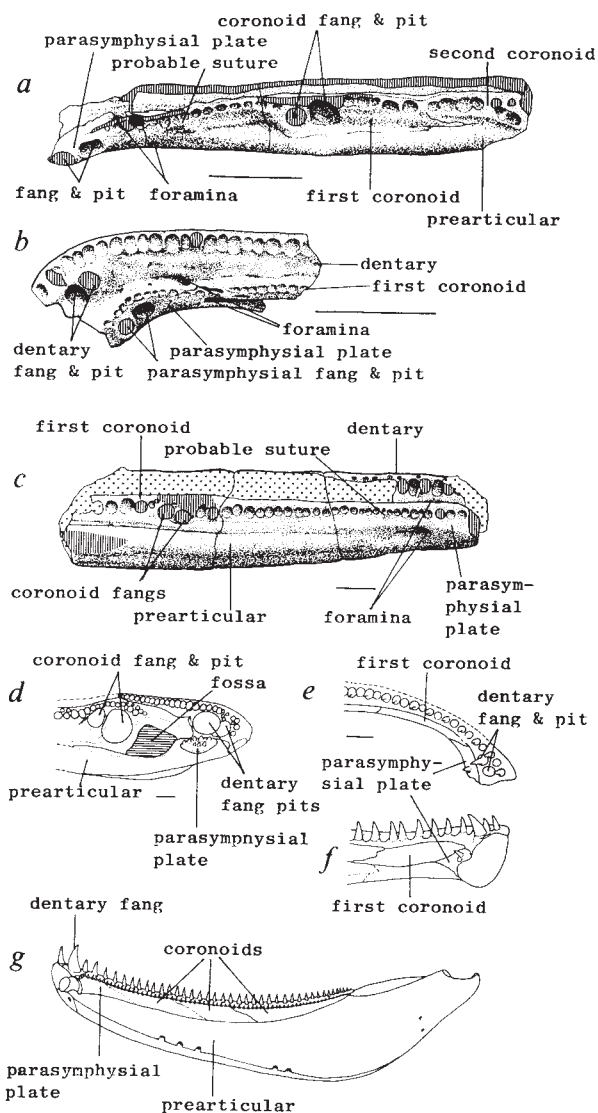


FIG. 3 a, Anterior part of right mandible from Scat Craig, Natural History Museum, London, specimen BMNH P62809. Dorsal view. This specimen lacks the dentary. Although most of the sutures are fused, the parasymphysial plate can be identified by comparison with b, and the parasymphysial/coronoid suture appears to be represented by a slight 'step' in the tooth row (compare with c). Hatching and line-width conventions of a and b as for Fig. 2. b, Holotype of *Obruchevichthys gracilis* Vorobyeva⁷, Palaeontological Institute, Moscow, specimen PIN 1491/52. Same orientation as a. In this specimen the dentary is preserved. c, Left mandibular fragment from Scat Craig, Royal Museum of Scotland, Edinburgh, specimen RMS G 1967.17.1. Mesiodorsal view. Hatching and line-width conventions as for Fig. 2; crosses, matrix. The cross-sections afforded by the fractures show that the coronoid and prearticular are broadly sutured together without exposure of the mentomandibular. Note the tiny marginal teeth lateral to the dentary tooth row; such teeth are absent in *Obruchevichthys*, but present in the Scat Craig premaxillae OUM D796 and BMNH P9776. The total length of this jaw can be estimated to be between 350 and 400 mm. d, *Panderichthys rhombolepis* (modified from ref. 30, parasymphysial plate after ref. 34), anterior part of left mandible in dorsal view. Horizontal hatching, exposed meckelian bone. e, f, Tetrapod from Westphalian A of Parrsboro, Nova Scotia (from ref. 33), anterior part of left mandible in dorsal and lateral views. The parasymphysial plate carries a fang and replacement pit. g, Right mandible of *Ichthyostega* (from ref. 2), mesial view. Coronoids and palatal bones both lack fangs in *Ichthyostega*, but the parasymphysial plate carries a continuation of the marginal coronoid tooth row. Note that the tetrapod jaws (e–g) lack the pair of foramina which 'bracket' the parasymphysial/coronoid suture in the Scat Craig mandibles, *Obruchevichthys* and *Metaxygnathus*. Scale bars, 10 mm.

the mandibles on the basis of tooth structure. They show that the skull was flat-snouted, as in panderichthyids and early tetrapods; unlike the skull of *Panderichthys* (my unpublished observation), they carry a tetrapod-like ornament of radially arranged, irregular pits and ridges.

The Scat Craig tibia represents the earliest known tetrapod-type hindlimb, the jaws belong to an animal apparently more closely related to tetrapods than *Panderichthys* or any other known fish, and the humerus possesses a suite of tetrapod characters. It is tempting to assume that they are parts of the same organism; their sizes are compatible, and the number of jaw fragments is large enough to suggest that the apparent absence of more than one tetrapod-like form is not due to sampling error. But although the tibia clearly belongs to a leg,

the humerus differs enough from the early tetrapod pattern to make it uncertain whether the appendage carried digits or a fin. At first sight the combination of two such extremities in one animal seems highly unlikely on functional grounds. If, however, tetrapod limbs evolved for aquatic rather than terrestrial locomotion, as recently suggested^{35,36}, such a morphology might be perfectly workable. Note that the forelimb of *Acanthostega*¹⁶ is more fish-like than the hindlimb and could probably not be brought into a weight-bearing position. Regrettably, the fragmentary nature of the Scat Craig material makes it impossible to determine whether the tibia and humerus really belong together, or with the jaws; it does show, however, that tetrapods or very tetrapod-like animals had appeared before the end of the Frasnian. □

Received 26 July; accepted 14 October 1991.

1. Säve-Söderbergh, G. *Meddr. Grönland* **94**, 1–107 (1932).
2. Jarvik, E. *Basic Structure and Evolution of Vertebrates* vol. 1 (Academic, London, 1980).
3. Lebedev, O. A. *Dokl. Akad. Nauk SSSR* **278**, 1407–1473 (1984).
4. Campbell, K. S. W. & Bell, M. W. *Alcheringa* **1**, 369–381 (1977).
5. Warren, J. W. & Wakefield, N. A. *Nature* **238**, 469–470 (1972).
6. Warren, A., Jupp, R. & Bolton, B. *Alcheringa* **10**, 183–186 (1986).
7. Vorobyeva, E. I. *Trudy paleont. Inst.* **163**, 1–239 (1977).
8. Schultze, H.-P. & Arsenault, M. *Palaeontology* **28**, 293–309 (1985).
9. Schultze, H.-P. *J. Morph.* (suppl. 1) 39–74 (1987).
10. Ahlberg, P. E. *Zool. J. Linn. Soc.* (in the press).
11. Vorobyeva, E. I. *Neues Jb. Geol. Paläont. Mh.* **1975**, 315–320 (1975).
12. Tarlo, L. B. *J. geol. Soc. Lond.* **117**, 193–213 (1961).
13. Miles, R. S. *Palaeontogr. Soc. (Monogr.)* **552**, 1–130 (1968).
14. House, M. R. et al. *Spec. Rep. Geol. Soc. Lond.* **8**, 1–110 (1977).
15. Harland, W. B. et al. *A Geologic Time Scale* (Cambridge University Press, 1982).
16. Coates, M. I. & Clack, J. A. *Nature* **347**, 66–69 (1990).
17. Gaffney, E. S. *Bull. Carnegie Mus. nat. Hist.* **13**, 92–105 (1979).
18. Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341–438 (1987).
19. Godfrey, S. J. *Phil. Trans. R. Soc. B323*, 75–133 (1989).
20. Holmes, R. *Phil. Trans. R. Soc. B306*, 431–527 (1984).
21. Panchen, A. L. *Phil. Trans. R. Soc. B309*, 505–568 (1985).

22. Smithson, T. R. *Zool. J. Linn. Soc.* **85**, 317–410 (1985).
23. Andrews, S. M. & Westoll, T. S. *Trans. R. Soc. Edinb.* **68**, 207–329 (1970).
24. Rackoff, J. S. in *The Terrestrial Environment and the Origin of Land Vertebrates* (ed. Panchen, A. L.) 255–292 (Academic, London, 1980).
25. Long, J. A. *Palaeontology* **30**, 839–852 (1987).
26. Andrews, S. M. & Westoll, T. S. *Trans. R. Soc. Edinb.* **68**, 391–489 (1970).
27. Ahlberg, P. E. *Zool. J. Linn. Soc.* **96**, 119–166 (1989).
28. Schultze, H.-P. *Palaeontogr. Ital.* **65**, 59–137 (1969).
29. Jessen, H. *Ark. Zool.* **18**(2), 305–389 (1966).
30. Gross, W. *Abh. preuss. Akad. Wiss. Math.-naturw. Kl.* **1941**, 1–51 (1941).
31. Smithson, T. R. *Palaeontology* **23**, 915–923 (1980).
32. Smithson, T. R. *Zool. J. Linn. Soc.* **76**, 29–90 (1982).
33. Godfrey, S. J. & Holmes, R. B. *Can. J. Earth Sci.* **26**, 1036–1040 (1989).
34. Vorobyeva, E. I. *Trudy. Paleont. Inst.* **104**, 1–108 (1962).
35. Edwards, J. L. *Am. Zool.* **29**, 235–254 (1989).
36. Coates, M. I. & Clack, J. A. *Nature* **352**, 234–235 (1991).

ACKNOWLEDGEMENTS. I thank P. L. Forey, S. M. Andrews and H. C. Ivimey-Cook for lending me Scat Craig specimens in their care, O. A. Lebedev for helpful comments on the material and access to the *Obruchevichthys* specimens in the Palaeontological Institute, Moscow, K. S. W. Campbell for allowing me to examine the only known specimen of *Metaxynghus* and J. A. Clack and M. I. Coates for access to the *Acanthostega* material, and for criticizing a previous version of this paper.

Primary structure and functional expression of a developmentally regulated skeletal muscle chloride channel

Klaus Steinmeyer, Christoph Ortland & Thomas J. Jentsch*

Centre for Molecular Neurobiology (ZMNH), Hamburg University, Martinistrasse 52, D-2000 Hamburg 20, Germany

SKELETAL muscle is unusual in that 70–85% of resting membrane conductance is carried by chloride ions¹. This conductance is essential for membrane-potential stability, as its block by 9-anthracene-carboxylic acid and other drugs causes myotonia^{2,3}. Fish electric organs are developmentally derived from skeletal muscle, suggesting that mammalian muscle may express a homologue of the *Torpedo marmorata* electroplax chloride channel^{4,5}. We have now cloned the complementary DNA encoding a rat skeletal muscle chloride channel by homology screening to the Cl[−] channel from *Torpedo*⁴ (Fig. 1a). It encodes a 994-amino-acid protein which is about 54% identical to the *Torpedo* channel and is predominantly expressed in skeletal muscle. Messenger RNA amounts in that tissue increase steeply in the first 3–4 weeks after birth, in parallel with the increase in muscle Cl[−] conductance⁶. Expression from cRNA in *Xenopus* oocytes leads to 9-anthracene-carboxylic acid-sensitive currents with time and voltage dependence typical for macroscopic muscle Cl[−] conductance. This and the functional destruction of this channel in mouse myotonia⁷ suggests that we have cloned the major skeletal muscle chloride channel.

We have called the cloned channel CIC-1 to distinguish it

from the *Torpedo* channel (CIC-0)⁴ and a recently cloned different mammalian Cl[−] channel, CIC-2 (A. Thiemann, S. Gründer, M. Pusch and T.J.J., manuscript in preparation). The predicted protein (994 amino acids, relative molecular mass 110,000 (M_r , 110K)) is larger than the *Torpedo* Cl[−] channel (~89K)⁴. This is due to amino- and carboxy-terminal extensions and to an insertion between domains D12 and D13. Hydropathy analysis⁸ (Fig. 1b) suggests a topology similar to the *Torpedo* Cl[−] channel. The greatest homology with the *Torpedo* protein is in putative membrane spans and adjacent regions. There is also considerable conservation of D13, which, by hydropathy, is a poor candidate for a transmembrane segment. This suggests that D13 is functionally important. Indeed, truncating the *Torpedo* channel between D12 and D13 renders it nonfunctional in the oocyte (K.S., C. Schmekal and T.J.J., unpublished observation).

The tissue distribution of CIC-1 was examined by northern analysis (Fig. 2a). A prominent band of about 4.5 kilobases (kb) was present in skeletal muscle. Faint bands of identical size were detected in kidney, liver, heart and in the smooth muscle cell line A10 (ref. 9). Thus, CIC-1 is predominantly, but probably not exclusively, expressed in skeletal muscle.

In rats, muscle Cl[−] conductance increases steeply during the first few weeks after birth⁶. CIC-1 mRNA amounts in muscle increase rapidly between days 1 and 30 after birth (Fig. 2b), suggesting a rise in channel density as the primary mechanism of postnatal conductance increase. Developmental changes in expression are also known for other muscle channels^{10–15}, possibly suggesting similar mechanisms of regulation.

Injection of CIC-1 cRNA into oocytes leads to functional expression of Cl[−] channels open under resting conditions, thereby clamping the oocyte membrane to the chloride equilibrium potential (−20 to −30 mV). Conductance is nearly linear between −80 and −20 mV, but decreases at more positive values (Fig. 3). At voltages more negative than about −80 mV, currents become progressively time-dependent. Hyperpolarizing voltage

* To whom correspondence should be addressed.

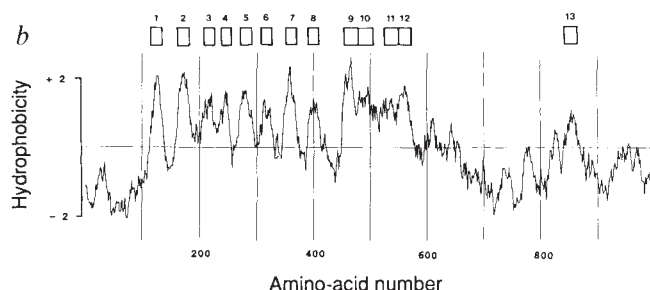


FIG. 1 Sequence and structural prediction for the rat skeletal muscle chloride channel (CIC-1). *a*, Nucleotide sequence of CIC-1 and alignment of its deduced amino-acid sequence (single-letter code) with that of the Cl^- channel from *T. marmorata* electroplax (CIC-0). The initiator methionine was assigned to the first ATG downstream of a stop codon in the frame. It is surrounded by sequences suitable for eukaryotic initiation of translation²². The first nucleotide and amino acid of the translation start are designated as position 1. Putative transmembrane domains D1 to D13 are underlined. Identical residues are boxed. Gaps are introduced to maximize alignment. *, Potential N-linked glycosylation sites. Compared with CIC-0, the site after D8 is conserved, though it was thought to be cytoplasmic. As in *Torpedo*, another site is found after D13 (but at a different position). An additional site exists between D12 and D13. A consensus phosphorylation site for kinase A (Ser 682) is present somewhat downstream compared with *Torpedo*. A striking feature of the segment between D12 and D13 is a highly negatively charged stretch (beginning at residue 709) of seven alternate Asp and Glu residues followed by a cluster of Pro residues. *b*, Hydrophobicity analysis of the rat skeletal muscle chloride channel protein. The mean hydrophobic index of a nonadecapeptide was calculated⁸ and plotted as a function of amino-acid number. Putative membrane-spanning domains are indicated by boxes and numbered from 1–13.

METHODS. An oligo (dT)-primed λ gt10 cDNA library from rat skeletal muscle (Clontech) was screened with radiolabelled *T. marmorata* Cl^- -channel clone 7134 (ref. 4) under reduced stringency (25% formamide, $5\times\text{SSC}$, $5\times$ Denhardt's and 0.1% SDS at 42 °C). Additional clones were obtained by rescreening under high-stringency conditions. The complete sequence of CIC-1 was obtained from four overlapping partial cDNA clones (clones λ m59, λ m49, λ m13 and λ m26). Both strands were completely sequenced with T7 DNA polymerase (Pharmacia). Partial sequence was also obtained from other independent clones. It is identical to the one shown, with the following exceptions: clone λ m30 has a T instead of a C at position 1,724, changing the amino acid from P to L; and λ m14 has a T instead of a C at position 1,055, changing an A to a V. The sequence has been deposited in the EMBL/Genbank database (accession number X62894).

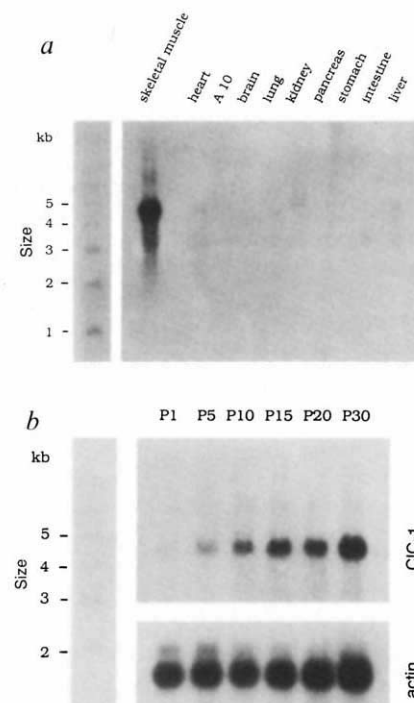


FIG. 2 Tissue distribution and developmental induction of chloride channel CIC-1. *a*, Expression of CIC-1 mRNA in different adult rat tissues as determined by northern analysis. Fragment lengths of a denatured ^{32}P -labelled DNA size standard (BRL) are shown on the left. *b*, Changes in CIC-1 mRNA amounts during skeletal muscle development. Northern analysis of mRNA isolated from rat skeletal muscle taken 1, 5, 10, 15, 20 and 30 days after birth (labelled P1 to P30) is shown. Lower part shows hybridization with γ -actin of the same blot to control for equal loading and absence of degradation.

METHODS. RNA was isolated from different rat tissues and A10 cells (a rat aorta smooth muscle cell line (ATCC CRL 1476)⁹ and enriched for poly(A) tracts. Poly(A)⁺ RNA ($10\text{ }\mu\text{g}$ per lane) was electrophoresed on 1% agarose gels containing formaldehyde, blotted and analysed using a full-length CIC-1 probe and standard techniques. Absence of degradation and equal amount of loading was tested by staining with ethidium bromide and control hybridization with γ -actin.

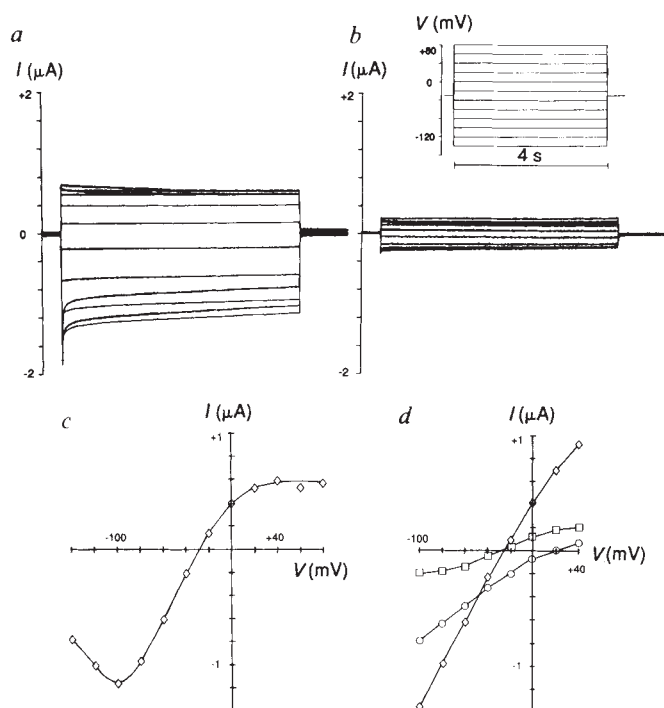


FIG. 3 Functional expression of CIC-1 chloride channel in *Xenopus* oocytes.

a, b, Standard two-electrode voltage clamp traces from oocytes previously injected with CIC-1 cRNA. *a*, Oocyte measured in normal saline (ND96)²³. *b*, The same oocyte measured about 15 min after bath application of 0.1 mM 9-AC. Inhibition by 9-AC needed several minutes to fully develop and was faster at higher concentrations. Inset (above *b*), voltage-clamp programme. From a holding potential of -30 mV , voltage was clamped for intervals of 4 s each to values between $+80$ and -140 mV . Superimposed traces in *a* and *b* represent currents injected into oocytes. *c*, Quasi-steady-state current-voltage relationship taken from experiment in *a* after a clamp time of 4 s. *d*, Current-voltage relationships of a different oocyte in normal saline ND96 (containing 103 mM Cl^-) ($\diamond$), about 15 min after application of 0.1 mM 9-AC ($\square$) and in low-chloride ND96 (7 mM Cl^- , 96 mM cyclamate^-) ($\circ$). **METHODS.** A full-length CIC-1 cDNA clone (F3) was assembled by ligating the 5' clone λ m59 at a *Ssp*I site (at base pair (bp) 205) to clone λ m49, followed by ligation to clone λ m13 at a *Bsp*II site (bp 1,010), and then to λ m26 (which extends beyond the stop codon at position 2,983) at the *Sal*I site (2,099 bp). As cRNA from that construct could not be expressed in oocytes, we used recombinant polymerase chain reaction (PCR)²⁴ to replace the muscle 5' untranslated sequence by 5' untranslated sequence derived from CIC-0 clone 7134 (ref. 4). This does not change any amino acid. The sequence of the fragment generated by PCR (ligated to F3 at the bp 205 *Ssp*I site) was fully verified. Capped cRNA was synthesized from this construct using T3 RNA polymerase after linearization of the construct. *Xenopus laevis* oocytes were prepared and injected²³ with 5–10 ng cRNA and measured 2 days later by two-electrode voltage clamp using pClamp software. The holding potential of -30 mV (chosen to be close to the resting voltage and chloride equilibrium potential) was constantly applied during the intervals (20 s) between individual voltage pulses.

steps initially elicit large currents, which then decay rapidly. The rate of deactivation increases with hyperpolarization. In this voltage range, steady-state currents actually decrease with hyperpolarization after passing through a maximum near -100 mV (Fig. 3c). Both observations agree remarkably well with studies on macroscopic skeletal muscle Cl^- conductance^{1,16}. Currents are predominantly carried by Cl^- , as partial replacement of extracellular Cl^- by impermeant cycamate⁻ reduces overall current and shifts the reversal potential ($I=0$) towards the new Cl^- -equilibrium potential (Fig. 3d). Further, conductance is $>80\%$ inhibited by 0.1 mM 9-anthracene-carboxylic acid (9-AC) (Fig. 3c, d), a Cl^- -channel inhibitor. Similar observations were made with macroscopic muscle Cl^- conductance, where application of 9-AC elicits myotonia^{2,3}.

There are several functional differences between this and the *Torpedo* channel, one being the sensitivity of 9-AC (CIC-0 is inhibited $<50\%$ by 2 mM 9-AC (T.J.J. and G. Schwarz, unpublished results). Although a decrease in open-probability at nega-

tive voltages, as observed with CIC-0, may also explain the current decrease with hyperpolarization for CIC-1, probably the most conspicuous difference is the lack of slow channel activation by hyperpolarization observed with CIC-0 (ref. 4). For CIC-0, this reflects a slow opening of a gate operating on both protochannels of the double-barrelled channel^{5,17-19}. Whether this implies that CIC-1 has no double-barrelled structure remains to be elucidated in single-channel studies. This is especially important as most patch-clamp studies on muscle Cl^- channels were done on undifferentiated myotubes^{20,21} whereas there are no single-channel data on the major Cl^- channel from intact differentiated muscle cells.

Thus the muscle Cl^- channel CIC-1, although in some regions highly homologous to the *Torpedo* channel CIC-0, has distinct electrophysiological properties. This channel is rather specifically expressed in skeletal muscle and probably provides the major Cl^- conductance in that tissue. Its importance for muscle function is best illustrated by the fact that its destruction in mouse mutants leads to myotonia⁷. □

Received 31 July; accepted 27 September 1991.

- Bretag, A. H. *Physiol. Rev.* **67**, 618-724 (1987).
- Rüdel, R. & Lehmann-Horn, F. *Physiol. Rev.* **65**, 310-356 (1985).
- Bryant, S. H. & Morales-Aguilera, A. *J. Physiol., Lond.* **219**, 367-383 (1971).
- Jentsch, T. J., Steinmeyer, K. & Schwarz, G. *Nature* **348**, 510-514 (1990).
- Miller, C. & Richard, E. A. in *Chloride Channels and Carriers in Nerve, Muscle, and Glial Cells* (eds Alvarez-Leefmans, F. J. & Russel, J. M.) 383-405 (Plenum, New York, 1990).
- Conte Camerino, D., De Luca, A., Mambriani, M. & Vrbová, G. *Pflügers Arch.* **413**, 568-570 (1989).
- Steinmeyer, K. et al. *Nature* **354**, 304-308 (1991).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
- Kimms, B. W. & Brandt, B. L. *Exp. Cell Res.* **98**, 349-366 (1976).
- Cooperman, S. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8721-8725 (1987).
- Trimmer, J. S. et al. *Neuron* **3**, 33-49 (1989).
- Trimmer, J. S., Cooperman, S. S., Agnew, W. S. & Mandel, G. *Dev. Biol.* **142**, 360-367 (1990).
- Kallen, R. L. et al. *Neuron* **4**, 233-242 (1990).

- Mishina, M. et al. *Nature* **321**, 406-411 (1986).
- Witzemann, V., Barg, B., Criado, M., Stein, E. & Sakmann, B. *FEBS Lett.* **242**, 419-424 (1989).
- Palade, P. T. & Barchi, R. L. *J. gen. Physiol.* **69**, 325-342 (1977).
- Miller, C. *Phil. Trans. R. Soc. B* **299**, 401-411 (1982).
- Miller, C. & White, M. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2772-2775 (1984).
- Bauer, C. K., Steinmeyer, K., Schwarz, J. R. & Jentsch, T. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Blatz, A. L. & Magleby, K. L. *Biophys. J.* **43**, 237-241 (1983).
- Blatz, A. L. & Magleby, K. L. *Biophys. J.* **47**, 119-123 (1985).
- Kozak, M. *Nucleic Acids Res.* **12**, 857-872 (1984).
- Colman, A. in *Transcription and Translation* (eds Hames, B. D. & Higgins, S. J.) 271-302 (IRL, Oxford, 1984).
- Higuchi, R. in *PCR Technology* (ed. Erlich, H. A.) 61-70 (Stockton, New York, 1989).

ACKNOWLEDGEMENTS. We thank C. Schmekel for technical assistance. This work is supported, in part, by the BMFT, the US Cystic Fibrosis Foundation, the US Muscular Dystrophy Association and the Deutsche Forschungsgemeinschaft.

Inactivation of muscle chloride channel by transposon insertion in myotonic mice

Klaus Steinmeyer*, Rainer Klocke†, Christoph Ortland*, Monika Gronemeier†, Harald Jockusch†, Stefan Gründer* & Thomas J. Jentsch*‡

* Centre for Molecular Neurobiology (ZMNH), Hamburg University, Martinistrasse 52, D-2000 Hamburg 20, Germany

† Developmental Biology Unit, W7, Bielefeld University, PO Box 8640, D-4800 Bielefeld 1, Germany

MYOTONIA (stiffness and impaired relaxation of skeletal muscle) is a symptom of several diseases caused by repetitive firing of action potentials in muscle membranes¹. Purely myotonic human diseases are dominant myotonia congenita (Thomsen) and recessive generalized myotonia (Becker), whereas myotonic dystrophy is a systemic disease. Muscle hyperexcitability was attributed to defects in sodium channels^{2,3} and/or to a decrease in chloride conductance (in Becker's myotonia⁴ and in genetic animal models⁵⁻¹⁰). Experimental blockage of Cl^- conductance (normally 70-85% of resting conductance in muscle¹¹) in fact elicits myotonia^{1,9}. ADR (ref. 12) mice are a realistic animal model^{5-7,12-18} for recessive autosomal myotonia. In addition to Cl^- conductance⁵, many other parameters^{6,12,16} are changed in muscles of homozygous animals. We have now cloned the major mammalian skeletal muscle chloride channel (CIC-1)¹⁹. Here we

report that in ADR mice a transposon of the *ETn* family²⁰⁻²³ has inserted into the corresponding gene, destroying its coding potential for several membrane-spanning domains. Together with the lack of recombination between the *Cic-1* gene and the *adr* locus, this strongly suggests a lack of functional chloride channels as the primary cause of mouse myotonia.

We first investigated whether in myotonic mice changes in the muscle chloride channel gene can be detected by genomic Southern analysis (Fig. 1). Using a rat muscle Cl^- channel (CIC-1)¹⁹ probe, aberrant fragments were indeed found in ADR (*adr/adr*) mice with several restriction nucleases, and with one enzyme in myotonic mice carrying the allelic mutation *adr*^{mt} (ref. 24). No rearrangement was found with the *adr*^K (ref. 25) allele. With ADR, aberrant fragments were always larger than the wild-type fragment, suggesting an insertional mutation. A total of 55 ADR mice were tested. In each of these the 6.6-kilobase (kb) *EcoRI* fragment was replaced by the 10.5-kb fragment, which was never found in any homozygous wild-type laboratory mouse. In proven heterozygous A2G mice, both the 6.6- and the 10.5-kb fragments were present.

Northern analysis was used to examine Cl^- channel messenger RNAs in various myotonic mouse strains. A probe from the 5' end of CIC-1 detected a roughly 4.5-kb message in skeletal muscle of both normal mice and myotonic mice homozygous for *adr*^{mt} and *adr*^K, whereas several bands were apparent with *adr/adr* mice (a roughly 7-8-kb doublet, and another doublet at about 1.6-2.0 kb) (Fig. 2a, d). Heterozygous (phenotypically normal) mice (*adr/+*) additionally had normal transcripts (about 4.5 kb), which were missing in homozygous (*adr/adr*) muscle. The sizes of the small mRNAs (1.6-2.0 kb) are insufficient to encode a functional Cl^- channel^{19,26}. To examine the *adr* mutation in detail, a complementary DNA library from (*adr/adr*) skeletal muscle was screened with CIC-1 cDNAs¹⁹. All 11 clones isolated were homologous to the rat muscle Cl^- -channel cDNA 5' to the sequence encoding the ninth putative

‡ To whom correspondence should be addressed.

transmembrane domain (D9), but no homologous sequence 3' to D9 was found. Four cDNAs extended into D9 (Fig. 3), but diverged (to two different sequences) after the triplet encoding lysine (amino-acid 467 in the rat¹⁹). This position corresponds exactly to an exon-intron boundary in the rat channel, suggesting that in ADR mice this exon is spliced to sequences unrelated to the Cl⁻ channel. Database searches showed that these are nearly identical with known members of the *Etn* family of mouse transposons²⁰⁻²³. High homology exists to an *Etn* sequence which had inserted into the switch region of immunoglobulins²³. In other regions, where no sequence information for this particular transposon is available, we found near identity to a fully sequenced genomic clone (MG1) of another *Etn* transposon²². We assume that a single transposon, which displays high homology to the first member and continues as MG1, has inserted into the Cl⁻-channel gene.

Among the four cDNAs examined, three different splice variants were found (Fig. 3). Two clones (*cDNA* 2 and 3), using the polyadenylation site in one of the long terminal repeats (LTR) of *Etn*, terminated with poly(A) tails. The predicted mRNA lengths are about 1.6 kb and 1.8 kb, corresponding to the small doublet in northern blots (Fig. 2). The *cDNA* 1 originated from splicing into *Etn* shortly downstream of the 5' LTR polyadenylation site, and the mRNA may continue until the 3' LTR polyadenylation signal. This could explain the long (roughly 7 kb) transcripts detected in northern blots. Thus insertion of a transposon (probably into the intron in D9) leads to abnormal splice products in which the sequence 3' of D9 is

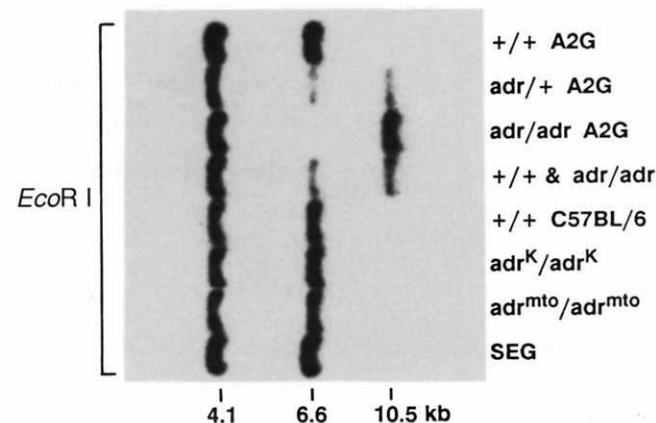


FIG. 1 Southern blot analysis of chloride channel gene ClC-1 in different mouse strains and mutants. DNAs were from normal and myotonic mice for which the genotypes and genetic backgrounds are indicated (SEG = *Mus spretus*). *EcoRI* digestion reveals aberrant restriction fragments in ADR mice and *adr/+* heterozygotes. In the lane designated '+/+ & *adr/adr*', a 1:1 mixture of A2G +/+ and A2G *adr/adr* DNAs was analysed. Aberrant restriction fragments were detected in ADR mice with the following nucleases (fragment-length, mutant (kb)/fragment length, normal (kb)): *BamHI* (13/11.5), *EcoRI* (10.5/6.6), *PstI* (7.4/5.8) and *SspI* (4.5/3.5). No differences were found with *BclI*, *BglII*, *EcoRV*, *HincII*, *HindIII*, *HpaI*, *MspI*, *PvuII*, *RsaI*, *ScaI*, *SphI*, *SstI*, *StuI*, *TaqI* or *XbaI*. In myotonic mice homozygous for the *adrmt0* or *adrK* mutations, no abnormal fragments were found with any of these nucleases, but *HincII* detected an abnormal fragment in *adrmt0*. Lengths of fragments are indicated.

METHODS. A2G mice carrying the recessive autosomal mutation, 'arrested development of righting response' (*adr*)¹², were obtained from R. L. Watts and D. L. Watts, Guy's Hospital, London. SWR/J mice carrying the allele *adrmt0* (ref. 24) were from the Jackson Laboratory, Bar Harbor, ME. The *adrmt0* gene was outbred into the more vigorous inbred strain C57BL/6J. The *adrK* breeders²⁵ (noninbred background) were obtained from P. Neumann, Harvard Medical School. DNAs were prepared from lungs. About 10 µg digested DNA was separated on a 0.8% agarose gel, blotted onto a nylon membrane and hybridized with a ³²P-labelled ClC-1 (ref. 19) probe extending from 970 to 1,881 bp of the rat sequence.

replaced by *Etn* sequences. To ascertain that the Cl⁻-channel sequence 3' to D9 is missing in all ADR transcripts, we probed the northern blot with a 3' cDNA. No signal was found with homozygous ADR animals, whereas the normal ~4.5-kb message was detected in normal mice, heterozygous *adr/+* animals and homozygous animals of other myotonic strains (Fig. 2b). An *Etn* probe detected transcripts only in *adr/adr* and *adr/+* RNA (Fig. 2c, e). They corresponded exactly to those recognized by the 5' Cl⁻-channel probe. Thus, all detectable Cl⁻-channel mRNA is functionally destroyed in ADR mice.

Using interspecies backcrosses, we mapped the *Clc-1* locus on mouse chromosome 6 between marker genes *Tcrb* and *Hox-1.1*, 1.3 centiMorgans (cM) ± 1.0 cM from *Tcrb* (Fig. 4). This is

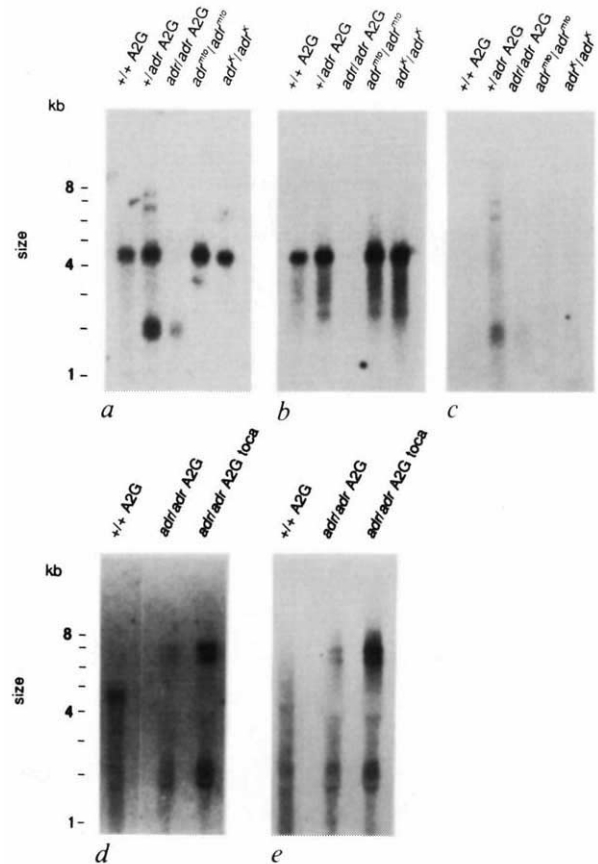


FIG. 2 Northern analysis of muscle Cl⁻-channel RNAs in normal and myotonic mice. Skeletal muscle RNA from the mouse strains indicated were analysed by probing blots with a rat skeletal muscle chloride channel (ClC-1) probe 5' to D9 (a, d), a ClC-1 probe 3' to D9 (b) and a probe derived from *Etn* sequences (c, e). a-c derive from the same blot hybridized sequentially with different probes. The mRNAs analysed on the blot used for d and e is from individuals different from those of a-c.

METHODS. RNA was isolated from leg skeletal muscle of adult mice of the indicated strains. (*adr/adr*) toca indicates homozygous ADR mice treated for 29 days with 1.5 mg tocanide per g dry food, which partially reverts the myotonic phenotype^{13,16}. About 3 µg poly (A)⁺ RNA (a-c) or 30 µg total RNA (d, e) per lane were electrophoresed on denaturing agarose gels, blotted onto nylon membranes and probed with ³²P-labelled cDNA probes. The 5' probe extends from -83 to 1,137 bp, and the 3' probe from 1,542 to 3,523 bp of the rat ClC-1 sequence¹⁹. The *Etn* probe is derived from clone *cDNA*2, extending from the *XbaI* site immediately after the first splice site (equivalent to position 268 of *Etn* A) to the poly(A) tail. Equal loading and absence of degradation was checked by ethidium bromide staining, but control hybridization with a γ -actin cDNA gave weaker signals for lanes 1 and 3 (a-c) and lanes 2 (d, e); the amounts of transcripts in these lanes may thus be underestimated by a factor of about 3.

In adult skeletal muscle, Cl^- channels provide the major resting conductance and stabilize its plasma-membrane voltage¹¹. Shortly after birth, K^+ conductance is dominant in rodent muscle²⁷. It decreases during the following 2–4 weeks, whereas Cl^- conductance increases to its predominant value²⁷ in parallel with *ClC-1* mRNA amounts¹⁹. Indeed, myotonia in ADR mice

ADR mice have previously been thoroughly investigated biophysically and biochemically^{5,12-17}. In addition to a reduced Cl^- conductance⁵ (now known to be the primary defect), various

b Chloride channel

ETn transposon

exon intron IR LTR IR LTR IR LTR IR

--ATG--

cDNA 1

cDNA 2

cDNA 3

AAAAAA

AAAAAA

AAAAAA

100 bp

other parameters^{6,12,16} were altered. Some changes were shown to be secondary to muscle hyperactivity¹⁶. In addition to a mutated Cl^- channel, several other mechanisms have been considered to cause myotonia^{6,15}. This work demonstrates the pleiotropic effects a defect in a single ion channel can have.

Thus *ETn* insertion has been shown to cause inheritable disease. *ETn* transposons were first identified as transcripts in teratocarcinoma cells²⁰ and during early embryonic development²¹. In these cases, a single polyadenylated RNA of ≥ 5 kb is transcribed, starting from a promoter in *ETn*. Here, of course, the Cl^- -channel promoter is used. Interestingly, differential splicing also occurs in *ETn*, suggesting an important influence of *cis* sequences on the use of potential splice sites. Although Cl^- -channel inactivation by *ETn* insertion is unique for the *adr* mutation, it is clear that in *adr^{mto}* and *adr^K*, whose mutation is allelic to *adr*, the defect also resides in the Cl^- channel and will be due to other types of mutations (such as point mutations).

ADR mice are a realistic animal model^{6,7} for human recessive generalised myotonia, which displays the same mode of inheritance and a reduced muscle Cl^- conductance⁴, though changes in Na^+ channels were also described³. Both changes in Cl^- conductance²⁸ and Na^+ channels² were also observed with myotonia congenita, a dominant disease. If these diseases were due to Cl^- -channel defects, they would be expected to map to

FIG. 3 Mechanism of Cl^- -channel inactivation in ADR mice. *a*, Sequence analysis of splice variants. Sequences of three ADR cDNAs (cDNA 1–3) are compared with the rat muscle Cl^- -channel sequence (rat mus)¹⁹ to a rat genomic clone, and to two members of the *ETn* family of transposable elements (*ETn* Ig, transposon encountered in the switch region of an immunoglobulin gene²³, GeneBank accession number M29266; *ETn* A, sequence of genomic clone MG1 (ref. 22), GeneBank accession No. M16478). Where appropriate, the corresponding protein translation is shown (the rat channel is compared at the protein level only). Putative transmembrane spans are indicated (D7 to D9). Only differences in sequence are shown; dashes indicate gaps, and points identities. The polyadenylation site in the LTR is underlined; splice consensus sequences are highlighted by lower case letters; IR, the inverted repeat; numbers, equivalent base pairs in *ETn* A. Predicted *adr* translation products are $>95\%$ identical to the rat muscle Cl^- -channel protein until it reaches the exon–intron boundary in D9. In cDNA 1 and 2 the exon is spliced into the 5' LTR of an *ETn* transposon at the same position (268). cDNA 2 leaves 167 bases downstream by splicing into the 3' LTR (where *ETn* Ig and *ETn* A are nearly identical). It continues until it is polyadenylated. cDNA 1 continues to be highly homologous to *ETn* Ig over the entire range where sequence information for this transposon is available. More downstream, sequence identity to *ETn* A is found. cDNA 3 splices directly into the (3' or 5') LTR and is polyadenylated. cDNA 2 and 3 splice into different reading frames of *ETn*, resulting in different translations. In all three splice variants, stop codons are encountered. *b*, Model for Cl^- -channel inactivation in ADR mice. Top, proposed genomic structure with *ETn* transposon (of about 5.5 kb total length) having inserted somewhere into the Cl^- -channel intron disrupting the sequence encoding putative transmembrane domain D9. IR, inverted repeats; LTR, long terminal repeat; AATAAA, polyadenylation site. Below, schematic representation of splice variants cDNA 1–3. As sequences in both LTRs are nearly identical, two splice mechanisms cannot be distinguished for cDNA 3 (one is depicted by a dashed line).

METHODS. A cDNA library in λ gt10 was constructed (clontech) by oligo (dT) and random primed synthesis starting from 7 μg poly (A)⁺ RNA from ADR skeletal muscle. About 2×10^6 primary recombinants were screened with a mixture of radiolabelled rat muscle Cl^- -channel cDNAs¹⁹ covering most of the coding sequence (extending from bp 432 to 632, and from bp 982 to 2287). About 60 positive clones were identified, 11 of which were isolated and sequenced from both ends. Four clones (named cDNA 1–4) extending into domain D9 were sequenced using T7 DNA polymerase. cDNA 3 and 4 represented identical splice variants. Only relevant regions were sequenced. The clones began at the following 5' positions (numbers according to the rat channel¹⁹): cDNA 1, 383 bp; cDNA 2, –37 bp; and cDNA 3, –65 bp. For genomic analysis, a rat genomic library in EMBL3 was screened using a rat Cl^- -channel cDNA¹⁹. The sequence shown was obtained from a *Hind*III fragment which was subcloned into pBluescript. The sequences have been deposited in the EMBL/GenBank database under accession numbers: X62895, X62896 and X62897.

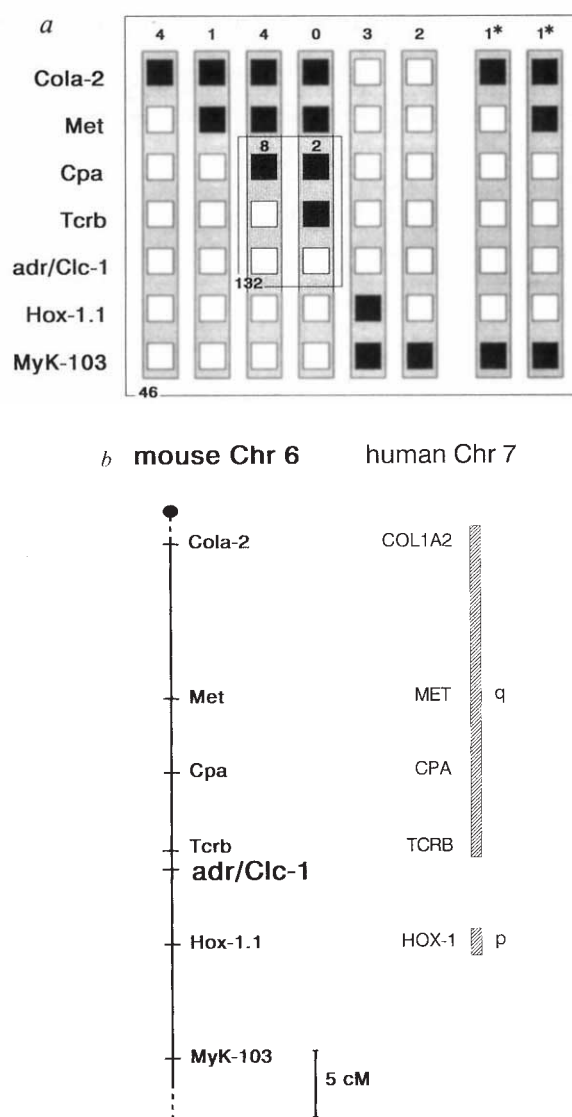


FIG. 4 Mapping of the *adr/Cic-1* gene on mouse chromosome 6 using interspecies backcross (BC). *a*, Schematic representation of experimental results used to map the locus. Recombinant haplotypes (indicated by vertical bars) of *Mus musculus domesticus* (Mmd) $\times$ *Mus spretus* (SEG/1) BC individuals are shown. Marker genes tested are given at left. Observed numbers of recombinants are indicated at the top of each haplotype. Empty squares indicate that the marker gene allele originated from the same species as the *adr/Cic-1* allele (no recombination), filled squares indicate a different origin (recombination). Asterisks indicate double recombinants. Frames indicate sets of marker genes tested, with inserted numbers indicating total numbers of backcross individuals on which these tests were performed, that is all markers shown at left were tested on 46 individuals, whereas 3 markers (*Cpa*, *Tcrb*, *adr/Cic-1*) were tested on 132 individuals, including the 46 ADR animals in the outer frame. Outer frame shows a result of a (Mmd A2G *adr*/+ $\times$ SEG/1 +/+) $\times$ Mmd A2G *adr*/+ BC, in which the ADR phenotype was used to indicate the *Mus musculus domesticus* origin of the F_1 -derived allele of *adr*. Inner frame represents results from the combined scores of ADR phenotypes and *Cic-1* alleles (which now became possible), and are in part (72 individuals) derived from a (C57BL/6J $\times$ SEG/1) $\times$ C57BL/6J BC (see ref. 29). *b*, Genetic map of proximal mouse chromosome 6, as based on recombination data from 150 BC individuals part of which are shown in *a*, compared with regions of human chromosome 7 to show conserved synteny (from ref. 18).

METHODS. DNAs from BC individuals were isolated and analysed by Southern blotting as described in Fig. 1. Mapping was done previously on a limited number of individuals using the myotonic phenotype to identify the *adr* locus¹⁸; it could now be extended exploiting species differences in *Xba*I fragments detected by the *Cic-1* probe. Marker gene probes: *Cola-2*, probe NJ3 3.2, human³⁰; *Met*, pmtdD, human³¹; *Cpa*, murine³²; *Tcrb*, human³³; *Hox-1.1*, pHox1-B2.8 murine³⁴; *Myk-103*, Bgl 1.6 murine³⁵.

human chromosome 7 (Fig. 4). It is now possible to address these issues directly. □

Received 31 July; accepted 27 September 1991.

- Rüdel, R. & Lehmann-Horn, F. *Physiol. Rev.* **65**, 310–356 (1985).
- Iaizzo, P. A. et al. *Neuromusc. Disord.* **1**, 47–53 (1991).
- Frank, C. et al. *Muscle Nerve* **14**, 762–770 (1991).
- Rüdel, R., Ricker, K. & Lehmann-Horn, F. *Muscle Nerve* **11**, 202–211 (1988).
- Mehrke, G., Brinkmeier, H., & Jockusch, H. *Muscle Nerve* **11**, 440–446 (1988).
- Jockusch, H. in *The Dynamic State of Muscle Fibers* (ed. Pette, D.) 429–443 (Walter de Gruyter, New York, 1990).
- Rüdel, R. *Trends Neurosci.* **13**, 1–3 (1990).
- Lipicky, R. J. & Bryant, S. H. *J. gen. Physiol.* **50**, 89–111 (1966).
- Bryant, S. H. & Morales-Aguilera, A. *J. Physiol., Lond.* **219**, 367–383 (1971).
- Adrian, R. H. & Bryant, S. H. *J. Physiol., Lond.* **240**, 505–515 (1974).
- Bretag, A. H. *Physiol. Rev.* **67**, 618–724 (1987).
- Watkins, W. J. & Watt, D. C. *Lab. Anim.* **18**, 1–6 (1984).
- Reininghaus, J., Fuchtbauer, E.-M., Bertram, K. & Jockusch, H. *Muscle Nerve* **11**, 433–439 (1988).
- Fuchtbauer, E.-M., Reininghaus, J. & Jockusch, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3880–3884 (1988).
- Brinkmeier, H. & Jockusch, H. *Biochem. biophys. Res. Commun.* **148**, 1383–1389 (1987).
- Jockusch, H., Reininghaus, J., Stühfauth, I. & Zippel, M. *Eur. J. Biochem.* **171**, 101–105 (1988).

- Költgen, D., Brinkmeier, H. & Jockusch, H. *Muscle Nerve* **14**, 775–780 (1991).
- Jockusch, H., Schenk, S. & Gronemeier, M. *Mouse Genome* **86**, 216 (1990).
- Steinmeyer, K., Ortland, C. & Jentsch, T. *J. Nature* **354**, 301–304 (1991).
- Brület, P., Kaghad, M., Xu, Y.-S., Croissant, O. & Jacob, F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5641–5645 (1983).
- Brület, P., Condamine, H. & Jacob, F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2054–2058 (1985).
- Sonigo, P. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3768–3771 (1987).
- Shell, B., Szurek, P. & Dunnick, W. *Molec. cell. Biol.* **7**, 1364–1370 (1987).
- Heller, A. H., Eicher, E. M., Hallett, M. & Sidman, R. L. *J. Neurosci.* **2**, 924–933 (1982).
- Neumann, P. & Weber, T. *Mouse News Lett.* **83**, 157 (1989).
- Jentsch, T. J., Steinmeyer, K. & Schwarz, G. *Nature* **348**, 510–514 (1990).
- Conte Camerino, D., De Luca, A., Mambriani, M. & Vrbová, G. *Pflügers Arch.* **413**, 568–570 (1989).
- Lipicky, R. J., Bryant, S. H. & Salmon, J. H. *J. clin. Invest.* **50**, 2091–2103 (1971).
- Friedmann, J. M., Leibel, R. L. & Bahary, N. *Mammalian Genome* **1**, 130–144 (1991).
- Meyers, J. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3516–3520 (1981).
- Dean, M. et al. *Nature* **318**, 385–388 (1985).
- Quinto, C. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 31–35 (1982).
- Yanagi, Y. et al. *Nature* **308**, 145–149 (1984).
- Bucan, M. et al. *EMBO J.* **5**, 2899–2905 (1986).
- Wilkie, T. M. & Palmiter, R. D. *Molec. cell. Biol.* **7**, 1646–1655 (1987).

ACKNOWLEDGEMENTS. We thank C. Schmekal for technical assistance, K. Kaupmann and S. Schenk for interspecies hybrids and M. Koch, F. Ramirez, J. H. Friedmann, A. Püschel, R. B. Palmiter and L. C. Tsui for marker gene probes. Work in T.J.J.'s laboratory is supported by the BMFT, the US Cystic Fibrosis Foundation, the US Muscular Dystrophy Association and the DFG, and in H.J.'s laboratory by the DFG.

Peripheral deletion of self-reactive B cells

David M. Russell*, Zlatko Dembić†, Grant Morahan‡, J. F. A. P. Miller‡, Kurt Bürki§ & David Nemazee*||¶

* Division of Basic Sciences, Department of Pediatrics, National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver, Colorado 80206, USA

† Department of Molecular Biology, Hoffmann-La Roche, CH-4005 Basel, Switzerland

‡ Walter and Eliza Hall Institute of Medical Research, The Royal Melbourne Hospital, Victoria 3050, Australia

§ Department of Biotechnology, Sandoz, CH-4002 Basel, Switzerland

|| Department of Microbiology and Immunology, University of Colorado Health Sciences Center, Denver, Colorado 80206, USA

¶ To whom correspondence should be addressed

B LYMPHOCYTES are key participants in the immune response because of their specificity, their ability to take up and present antigens to T cells, and their capacity to differentiate into antibody-secreting cells. To limit reactivity to self antigens, autospic B cells can be functionally inactivated or deleted^{1–4}. Developing B cells that react with membrane antigens expressed in the bone

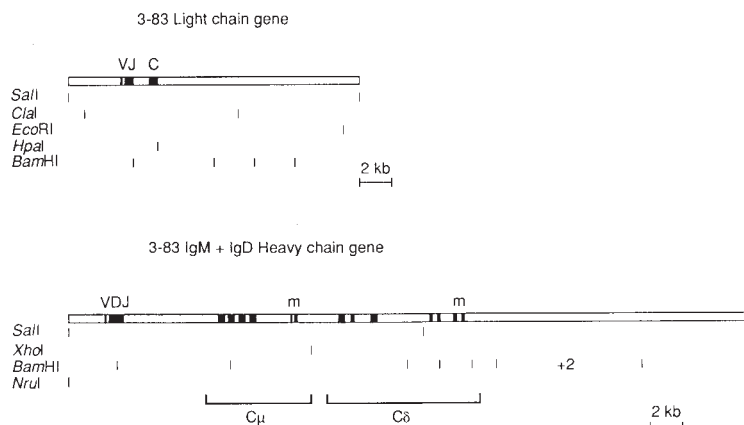
marrow are deleted from the peripheral lymphocyte pool^{4–6}. It is important to ascertain the fate of B cells that recognize membrane autoantigens expressed exclusively on peripheral tissues because B cells in the peripheral lymphoid organs are phenotypically and functionally distinct from bone-marrow B cells^{7–9}. Here we show that in immunoglobulin-transgenic mice, B cells specific for major histocompatibility complex class I antigen can be deleted if they encounter membrane-bound antigen at a post-bone-marrow stage of development. This deletion may be necessary to prevent organ-specific autoimmunity.

To test for peripheral deletion double-transgenic (Dbl-Tg) mice were produced by mating mice bearing the *MT-K^b* transgene, which targets expression of the membrane major histocompatibility complex (MHC) class I antigen *K^b* to the liver¹⁰, with 3-83 μ δ mice (Ig-Tg), which are transgenic for heavy and light chain immunoglobulin genes encoding anti-*K^b* antibody (see legend to Fig. 1). No messenger RNA or surface-protein expression of the *K^b* is detectable in lymphoid tissues of *MT-K^b* mice¹⁰. Both the Ig-Tg and *MT-K^b* transgenics were bred as hemizygous traits and (Ig-Tg × *MT-K^b*)F₁ littermates were analysed.

Double-transgenic Ig-Tg/*MT-K^b* mice deleted peripheral B cells. In Dbl-Tg lymph nodes there were >25-fold fewer cells bearing IgM(sIgM⁺-cells) compared with Ig-Tg or non-Tg controls and there were no detectable 3-83 idiotype-positive cells

FIG. 1 Maps of the DNA fragments used to generate 3-83 μ δ transgenic mice. Dark regions represent exons.

METHODS. Mice were produced using functional rearranged heavy- and light-chain genes encoding the 3-83 antibody. This antibody binds with moderate affinity to *K^k*, with weak affinity to *K^b*, and fails to bind to H-2^d cells²². Transgenic mice encoding the IgM form of 3-83 have been described^{4–6}. Self tolerance to *K^k* and *K^b* in the IgM-only 3-83 mice is mediated by deletion in H-2^d transgenic → H-2^d × H-2^k or H-2^d transgenic → H-2^d × H-2^k bone marrow chimaeras^{5,6}. To generate the 3-83 IgM plus IgD heavy-chain fragment, the 3-83 μ construct used previously⁴ was extended to include the complete *C δ* genomic locus. The 3-83 μ insert was liberated from its vector by partial digestion with *EcoRI* and cloned into the *EcoRI* site of λ EMBL3 to generate λ 241. A cosmid clone spanning the (DBA/2-derived) Ig-C μ and Ig-C δ regions was then isolated from a genomic library of the T-cell hybrid BDF1 16 (ref. 23), linearized with *XhoI*, which cuts at the site between *C μ* and *C δ* , and ligated together with the *VDJ/C μ* -containing *SalI/XhoI* fragment of λ 241 and with *SalI*-digested pNKL cosmid vector²⁴, resulting in a cosmid whose restriction map in the *C μ /C δ* region corresponds to the natural locus. The 42-kilobase (kb) insert was liberated from all but ~200 base pairs (bp) of the vector by *NruI* digestion. Light- and heavy-chain



gene fragments were isolated and transgenic mice produced as described⁴. Southern blotting and segregation analysis indicated that the 3-83 μ δ line has ~3–5 copies of the light- and heavy-chain genes co-integrated at a single chromosomal locus. m, Transmembrane exons.

FIG. 2 Elimination of autoreactive B cells in spleen and lymph nodes, but not in the bone marrow, of an Ig-Tg mouse bearing peripherally expressed antigen. Cells from the organs of 5-week-old transgenic and normal littermates were analysed by two-colour flow cytometry. *a*, Staining with anti-idiotypic and anti-IgM. Idiotype was detected with the 3-83 clonotype-specific rat monoclonal antibody 54.1, which recognizes an epitope created by the pairing of 3-83 heavy and light chains²⁰. *b*, Staining with anti-IgD and anti-IgM. Dbl-Tg, double transgenic (3-83 μ γ /MT-K^b); Ig-Tg, immunoglobulin transgenic; Non-Tg, non-transgenic. Cells were prepared as before and stained with monoclonal antibodies: anti-3-83 clonotype, 54.1/biotin; anti-IgM, DS1/FITC (ref. 25); anti-IgD^a, AMS 9.1/biotin (Pharmingen)²⁶. Biotin-conjugated antibodies were revealed with phycoerythrin (PE)-streptavidin stain (Becton-Dickinson). Purification and fluorochrome conjugation of the antibodies was as described²⁷. Flow cytometry analysis was on a Profile (Coulter) machine and data are presented using its density plot function. The percentages of cells in each quadrant, rounded to the nearest 0.1%, are indicated in the upper right corner of each plot. Data for bone marrow cells are gated by side scatter to exclude the predominant myeloid cell population from analysis. All mice used in this experiment were H-2^d and Igh-C^a at the endogenous loci. The genetic background of the mice is ~75% BALB/cJ, 6% B10.D2, 6% C57b16/J, 6% SJL/J, 6% DBA/2. The transgenic heavy chain μ and δ allotypes are type *a*. The results shown are representative of 4 experiments (Table 1). MT-K^b mice were indistinguishable from Non-Tg mice in these assays (not shown).

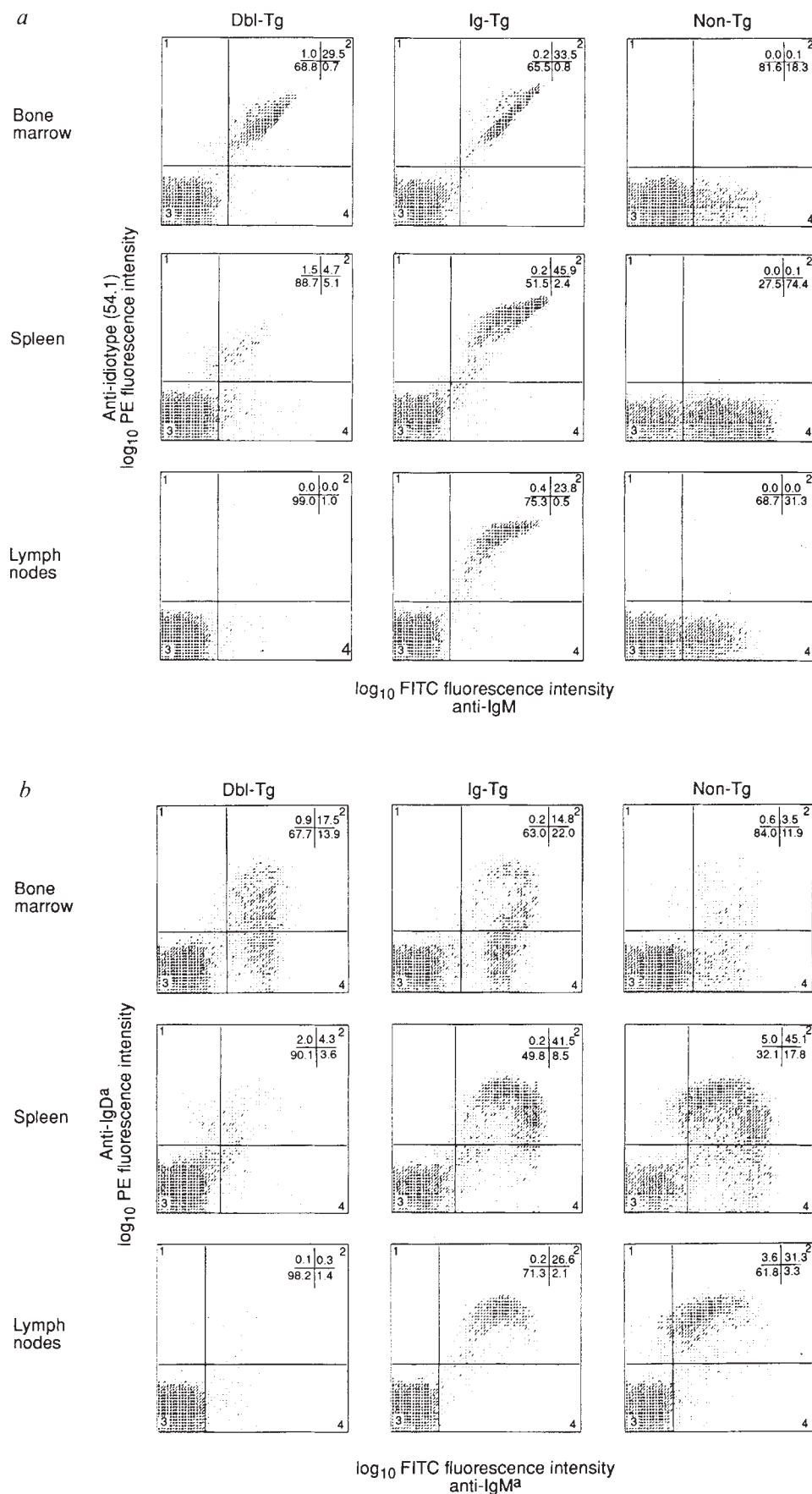


TABLE 1 Ability of peripherally expressed K^b antigen to induce B-cell tolerance as assayed by reduction in the amount of serum IgM idiotype and decrease in the number of peripheral B lymphocytes

Serum immunoglobulin*											
n	Type	IgM ($\mu\text{g ml}^{-1}$)	3-83 idiotype ($\mu\text{g ml}^{-1}$)								
2	Non-Tg	360 $\pm$ 90	<0.4								
5	Ig-Tg	98 $\pm$ 47	34 $\pm$ 16								
5	Dbl-Tg	103 $\pm$ 36	<0.4								
Peripheral B cells†											
n	Type	Total nucleated cells ($\times 10^{-6}$)		B220+ cells ($\times 10^{-6}$)		% IgM positive		% B220 positive			
		Spleen	Lymph nodes‡	Spleen	Lymph nodes	Spleen	Lymph nodes	Spleen	Lymph nodes	Spleen	Lymph nodes
3	Non-Tg	109 $\pm$ 16	67 $\pm$ 15	62	20	53.9 $\pm$ 14.6	28.2 $\pm$ 6.0	56.6 $\pm$ 12.1	30.0 $\pm$ 4.5		
4	Ig-Tg	105 $\pm$ 38	81 $\pm$ 21	43	20	44.4 $\pm$ 7.5	24.9 $\pm$ 5.1	41.1 $\pm$ 5.6	25.0 $\pm$ 3.6		
5	Dbl-Tg	77 $\pm$ 21	62 $\pm$ 20	12	0.7	9.1 $\pm$ 2.5	0.8 $\pm$ 0.4	11.4 $\pm$ 4.1	1.2 $\pm$ 0.6		

Immunoglobulin concentrations were measured using standard enzyme-linked immunosorbent assay (ELISA) techniques¹⁸. Polyvinylchloride plastic microtitre plates were coated with rat monoclonal antibodies specific for IgM (Ak15; ref. 19) or 3-83 idiotype (54.1). After washing and blocking, sera (diluted in PBS supplemented with 1% BSA) were incubated for 3 h at 25 °C. Bound immunoglobulin was detected by biotinylated rat anti-mouse IgM (Ak2; ref. 19) and developed using streptavidin-peroxidase (Sigma) followed by colorimetric substrate. Absorbance was measured on a BioRad model 2225 ELISA plate reader. Data were analysed using the Microplate Manager program. Standard curves for the assay were generated using a supernatant from Cos lin D1, an Sp2/0 myeloma transfected with genes encoding κ and μ chains of 3-83 (ref. 20). Immunofluorescence analysis was performed as described for Fig. 1. Anti-B220 (CD45R) antibody was FITC-RA3-3A1 (ref. 21).

* All data are presented as mean $\pm$ s.d. Mice were 6–8 weeks old.

† All data are presented as mean $\pm$ s.d. Mice were 4–8 weeks old. Data are from 4 independent experiments done on different days. Each experiment included one Ig-Tg mouse, at least one Dbl-Tg littermate and, in all but one experiment, a non-Tg littermate.

‡ Cells were pooled from the following lymph nodes: superficial inguinal, brachial, axillary, superficial cervical, mesenteric and popliteal.

(Fig. 2a; Table 1). This lack of IgM⁺ B cells in Dbl-Tg lymph nodes was not the result of sIgM downregulation because there was a similar >25-fold decrease in cells bearing the B-cell surface marker B220 (CD45R) (Table 1) and because >95% of the remaining cells were T cells bearing the $\alpha\beta$ receptor ($\alpha\beta$ -TCR⁺ T cells; data not shown). In Ig-Tg mice, >95% of bone marrow, spleen and lymph node B cells bore the 3-83 idiotype and these cells expressed IgM and IgD at levels as high as normal (Non-Tg) littermates (Fig. 2). In contrast to the deletion of idiotype-positive peripheral B cells, no deletion was evident in the bone marrow of Dbl-Tg mice, which had as many sIgM⁺/idiotype⁺ B cells as Ig-Tg littermates (Fig. 2a), as would be expected if

antigen was only encountered in the periphery. This is clearly different from 3-83 mice bearing K^b or K^k on the whole body, in which cells bearing a high density of sIgM are greatly depleted in the bone marrow (refs 4–6, and our unpublished results). Furthermore, the deletion is probably not the result of secreted or shed soluble K^b because of the tissue specificity of the deletion and because up to 200 ng ml⁻¹ of the high-affinity ligand K^k fails to tolerize 3-83 μ cells *in vivo*⁶.

Serum 3-83 idiotype levels were >80-fold lower in Dbl-Tg compared with Ig-Tg mice, as would be predicted if autoreactive B cells were eliminated (Table 1). Interestingly, the bulk of the spontaneously secreted IgM in these mice lacked the 3-83

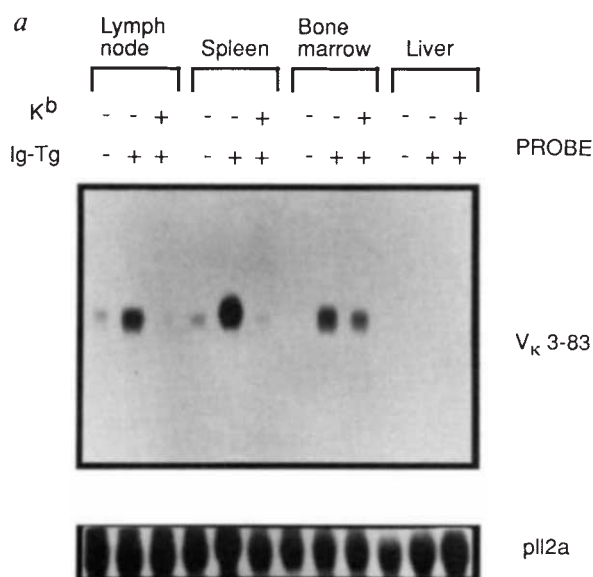
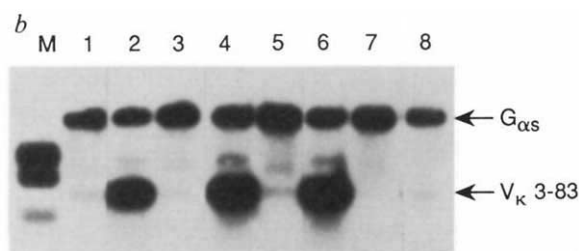


FIG. 3 RNA analysis of transgenic and non-transgenic mice. a, Northern blot analysis shows reduction in transgenic κ -light chain expression in spleen and lymph nodes of Dbl-Tg mice. After hybridization with V κ 3-83 probe⁴ and decay of the radioactive signal, the filter was hybridized with the MHC class I-specific pII2a probe²⁸. b, Polymerase chain reaction (PCR) detection of V κ 3-83 RNA in the livers of peripherally deleting mice. Lanes 1, 3, 5 and 7, liver; lanes 2, 4, 6 and 8, bone marrow. Lanes 1 and 2, mouse 2080



(Dbl-Tg); lanes 3 and 4, mouse 2081 (Dbl-Tg); lanes 5 and 6, mouse 2082 (Ig-Tg); lanes 7 and 8, Non-Tg littermate. Complementary DNA was synthesized and amplified with 15 cycles of PCR, run on a 1.2% agarose gel, transferred to a Zetaprobe membrane (BioRad) and hybridized with probes recognizing gene transcripts of 3-83 V κ or the GTP-binding protein G α_s (ref. 29). Lane M, marker DNA fragments of 587, 434 and 267 bp.

METHODS. RNA preparation and northern blotting were as described⁴. 5 μg RNA was loaded per lane in the 1% agarose formaldehyde gel. RNA samples were reverse-transcribed using a commercial kit (Superscript, BRL). PCR reactions were in a final volume of 30 μl and contained 1 μl reverse transcription reaction, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 20 $\mu\text{g ml}^{-1}$ gelatin, 0.3 mM each of dATP, dCTP, dGTP and dTTP, 1 μM of each oligonucleotide primer, and 1 U of *Taq* polymerase (Perkin-Elmer Cetus). Each cycle consisted of 30 s at 94 °C, 30 s at 62 °C and 1.5 min at 72 °C in a thermal cycler (Ericomp). The appropriate cDNA clones were used as hybridization probes after labelling with [α -³²P]dCTP. PCR primers were: 5'V κ 3-83 (79–101): 5'-CAGCTTCCTGCTAATCAGTGCC-3'; 3'J κ 2 (412–431): 5'-TGGTCCCCCTCCGAACGTG-3'; G α_s sense (309–336): 5'-ATTGAAACCA-TTGTGGCCCGCCATGAGC-3'; G α_s antisense (1,104–1,128): 5'-GAAGACACGG-CGGATGTCTCAGTGTC-3'.

idiotype and was present at a similar concentration in Dbl-Tg and Ig-Tg mice. Thus variant B cells that fail to produce the transgene-encoded idiotype are responsible for most of the circulating IgM in these mice.

Consistent with the interpretation that the autoreactive B cells are deleted, transgene 3-83 κ mRNA levels paralleled the levels of idiotype-positive B cells in the lymph nodes, spleen and bone marrow of Ig-Tg and Dbl-Tg mice (Fig. 3). In addition, 3-83 κ mRNA levels in the livers of Dbl-Tg mice were not elevated compared with Ig-Tg mice, indicating that there was no accumulation of autoreactive B cells in the antigen-bearing organs (Fig. 3a, b). This conclusion was also supported by histological examination of liver sections (not shown).

The Dbl-Tg mice allowed us to test *in vivo* the notion that expression of IgD by B cells renders them impervious to tolerance induction^{11,12}. The deleted cells in Dbl-Tg mice probably expressed IgD at the time they encountered antigen because IgD was expressed on a large fraction of the detectable B cells, including about half of the Dbl-Tg bone marrow B cells, which should not yet have encountered antigen (Fig. 2b). The absence of B cells in the lymph nodes of the Dbl-Tg mice indicates that these mice had a greatly depleted recirculating B-cell pool, and that most of the B cells present in the bone marrows and spleens of Dbl-Tg mice were probably newly formed. The remaining B cells in the spleens of Dbl-Tg mice had both IgD and idiotype. It is possible that these cells were generated by lymphopoiesis in the spleen or were in transit from the bone marrow and had not yet encountered K^b. In any case, there was a clear quantitative diminution of autoreactive cells that occurred outside the primary haematopoietic organ for B cells, the bone marrow. In addition, adoptive transfer of Ig-Tg lymph node and spleen B cells into 750-roentgen-irradiated K^b-bearing mice resulted in deletion (D.M.R. and D.N., unpublished results). These data are consistent with reports demonstrating that certain B-cell lymphomas undergo anti-Ig- κ -induced apoptosis *in vitro* whether or not sIgD is co-ligated with sIgM (refs 13, 14). On the other hand, IgD expression was probably not necessary for the observed deletion, because IgM-only 3-83 transgenic mice also delete autoreactive B cells in crosses with MT-K^b (ref. 6, and our unpublished data).

These data demonstrate B-cell clonal deletion to membrane autoantigens expressed in the periphery. They also indicate that B-cell deletion is unlikely to be mediated by a specialized antigen-presenting cell, such as a 'veto' cell¹⁵, because in MT-K^b mice the K^b antigen is only detectable on the surfaces of hepatocytes, exocrine pancreatic cells and kidney tubules¹⁰. On the other hand, we do not believe that all peripheral membrane autoantigens induce B-cell deletion as double-transgenic mice expressing K^b under the control of a keratin promoter fail to delete idiotype-positive B cells completely (D.N., B. Arnold and G. Hammerling, unpublished results). It is not yet clear whether K^b antigen density, location, availability, or timing of expression determine the differences between the MT-K^b/3-83 $\mu\delta$ and keratin-K^b/3-83 $\mu\delta$ Dbl-Tg mice.

Tolerance to many self proteins may rely on inactivation or deletion of antigen-specific T cells, B cells, or both. Because in some cases B cells are the limiting antigen-presenting cells for class II-restricted T-cell responses^{16,17}, deletion of autoreactive B cells should hinder both autoantibody formation and the presentation of autoantigens to class II-restricted T cells. Deletion of B cells in the periphery may be important in the prevention of tissue-specific autoimmune disease. □

Received 18 July; accepted 17 September 1991.

1. Nossal, G. J. V. *Science* **24**, 147-153 (1989).
2. Goodnow, C. C. *et al. Nature* **334**, 676-682 (1988).
3. Erikson, J. *et al. Nature* **349**, 331-334 (1991).
4. Nemazee, D. & Bürki, K. *Nature* **337**, 562-566 (1989).
5. Nemazee, D. & Bürki, K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8039-8043 (1989).
6. Nemazee, D. *et al. Immunol. Rev.* **122**, 117-132 (1991).
7. Raff, M. C. *et al. J. exp. Med.* **142**, 1052-1064 (1975).

8. Sidman, C. L. & Unanue, E. R. *Nature* **257**, 149-151 (1975).
9. Gu, H., Tarlington, D., Müller, W., Rajewsky, K. & Förster, I. *J. exp. Med.* **173**, 1357-1371 (1991).
10. Morahan, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 3782-3786 (1989).
11. Vitetta, E. S., Cambier, J. C., Ligler, F. S., Kettman, J. R. & Uhr, J. W. *J. exp. Med.* **146**, 1804-1810 (1977).
12. Scott, D. W., Layton, J. E. & Nossal, G. V. J. *J. exp. Med.* **146**, 1473-1483 (1977).
13. Tisch, R., Roifman, C. M. & Hozumi, N. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6914-6918 (1988).
14. Alés-Martinez, J. E., Warner, G. L. & Scott, D. W. *Proc. Natn. Acad. Sci. U.S.A.* **85**, 6919-6923 (1988).
15. Rammensee, H.-G., Bevan, M. J. & Fink, P. J. *Immun. Today* **6**, 41-43 (1985).
16. Chestnut, R. W. & Grey, H. M. *J. Immun.* **115**, 1075-1079 (1981).
17. Ron, Y. & Sprent, J. *J. Immun.* **128**, 2848-2856 (1987).
18. Nakamura, R. M., Volter, A. & Bidwell, D. E. in *Handbook of Experimental Immunology* 4th edn (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 26.1-27.20 (Blackwell, London, 1986).
19. Leptin, M. *et al. Eur. J. Immun.* **13**, 534-542 (1984).
20. Nemazee, D. A. in *Proc. EMBO Workshop on Tolerance* (eds Matzinger, P. *et al.*) Vol. 2 52-54 (Editiones Roche, Basel, 1987).
21. Coffman, R. L. & Weissman, I. L. *Nature* **289**, 681-683 (1981).
22. Ozato, K., Mayer, N. & Sachs, D. H. *J. Immun.* **124**, 533-540 (1980).
23. Dembič, Z. *et al. Nature* **320**, 232-238 (1986).
24. Grosfeld, F. G. *et al. Nucleic Acids Res.* **10**, 67145-6732? (1982).
25. Sieckman, D. *Fed. Proc.* **43**, 488 (1987).
26. Stall, A. M. & Loken, M. R. *J. Immun.* **132**, 787-795 (1984).
27. Hardy, R. R. in *Handbook of Experimental Immunology* 4th edn (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 97.1-97.17 (Blackwell, London, 1986).
28. Steinmetz, M. *et al. Cell* **25**, 125-134 (1981).
29. Jones, D. T. & Reed, R. R. *J. biol. Chem.* **262**, 13241-14249 (1987).

ACKNOWLEDGEMENTS. This work was supported by the NIH and the UCHSC Cancer Center. We thank G. Johnson for the G_{αs} primers, G. Rothhammer for technical assistance, W. Townsend for help with flow cytometry, and L. J. Wysocki, T. Finkel and R. Kubo for review of the manuscript.

Catalysis of guanine nucleotide exchange on the CDC42Hs protein by the *dbl* oncogene product

Matthew J. Hart*, Alessandra Eva†, Tony Evans‡, Stuart A. Aaronson† & Richard A. Cerione*

* Department of Biochemistry, Cellular and Molecular Biology, and Department of Pharmacology, Cornell University, Ithaca, New York 14853, USA

† Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892, USA

‡ Department of Cell Biology, Genentech Inc., 460 Pt San Bruno Boulevard, South San Francisco, California 94080, USA

THE superfamily of low molecular mass GTP-binding proteins, for which the *ras* proteins are prototypes, has been implicated in the regulation of diverse biological activities including protein trafficking, secretion, and cell growth and differentiation¹⁻³. One member of this family, CDC42Hs (originally referred to as Gp or G25K)^{4,5}, seems to be the human homologue of the *Saccharomyces cerevisiae* cell-division-cycle protein, CDC42Sc (refs 6-9). A second *S. cerevisiae* protein, CDC24 (ref. 10), which is known from complementation studies to act with CDC42Sc to regulate the development of normal cell shape and the selection of nonrandom budding sites in yeast, contains a region with sequence similarity to the *dbl* oncogene product¹¹⁻¹⁴. Here we show that *dbl* specifically catalyses the dissociation of GDP from CDC42Hs and thereby qualifies as a highly selective guanine nucleotide exchange factor for the GTP-binding protein. Although guanine nucleotide exchange activities have been previously described for other members of the Ras-related GTP-binding protein family¹⁵⁻¹⁷, this is the first demonstration, to our knowledge, of the involvement of a human oncogenic protein in catalysing exchange activity.

The *dbl* oncogene product¹¹⁻¹³ seemed to represent a candidate for a mammalian CDC42Hs-regulatory protein because a region of proto-*dbl* (residues 498-735) is 29% identical (~70% related when conservative amino-acid residues are taken into account) to the CDC24 protein (residues 163-400)¹⁴, which is known to participate in the same signalling pathway as CDC42Sc in *S. cerevisiae*. To examine whether *dbl* can regulate some aspect of the GTP-binding/GTPase cycle of CDC42Hs, we used a

FIG. 1 Expression of the *dbl* oncogene product in *S. frugiperda* cells. Western blot analyses of the immunoprecipitates (using anti-*dbl* protein antisera) of *S. frugiperda* lysates expressing the *dbl* oncogene product (depicted as p66^{dbl}, lane 2) and control lysates (lane 1).

METHODS. An expression vector (pC11(pZIP-neo)) containing the entire coding sequence of the *dbl* oncogene^{1,3} was digested with *Bam*HI and the released insert ligated to the *Bam*HI sites of the transfer vector PAC373. The resulting construct PAC373 *dbl* contains all but eight nucleotides of the baculovirus polyhedrin gene 5' untranslated sequences fused to the complete coding sequences of the *dbl* oncogene, including the first ATG codon. The *S. frugiperda* (Sf9) cells were infected by adsorption for 1 h with either wild-type baculovirus (control cells) or with *dbl* recombinant baculovirus and collected 60 h after infection. For analyses of the expression of p66^{dbl}, cells were lysed using a buffer containing 10 mM sodium phosphate, pH 7.4, 0.1 M NaCl, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 1 mM phenylmethylsulphonyl fluoride (PMSF), 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin and 10 µg ml⁻¹ pepstatin. Clarified cell lysates were immunoprecipitated using peptide antisera raised against amino acids 818–831 of proto-*dbl* and electrophoresed through an 8% polyacrylamide gel followed by immunoblotting with a peptide polyclonal rabbit antisera raised against amino acids 3–16, 592–606 and 818–831 of proto-*dbl* (ref. 12).

high-titred, recombinant baculovirus (free of wild type) to express the *dbl* oncogene product (from here on referred to as p66^{dbl}) in *S. frugiperda* cells¹⁸. Expression of the p66^{dbl} was verified by probing immunoprecipitates of lysates from *S. frugiperda* cells with rabbit proto-*dbl* antibodies raised against amino acids 818–831 of the proto-*dbl* sequence (Fig. 1). Neither the lysates containing p66^{dbl}, nor the control lysates from *S. frugiperda* cells, stimulated the GTPase activity of the human platelet CDC42 protein under conditions where a recently identified GTPase-activating protein (GAP) for CDC42Hs¹⁹ catalysed the hydrolysis of greater than 80% of the [γ -³²P]GTP bound

to the platelet CDC42Hs. Thus the *dbl* oncogene product is not a GAP for the CDC42Hs protein.

We next examined whether the lysates containing p66^{dbl} might influence the binding of [³H]GDP to the purified platelet CDC42Hs protein, either as a stimulator of GDP dissociation and the GDP-GTP exchange reaction or as a GDP dissociation inhibitor (GDI). Figure 2 shows that p66^{dbl} markedly catalyses the dissociation of [³H]GDP from the platelet CDC42Hs. The half-time for GDP dissociation in the presence of p66^{dbl} occurred within 2.5 min, and was essentially complete in ~20 min. By contrast, the control lysates caused little dissociation (≤20%)

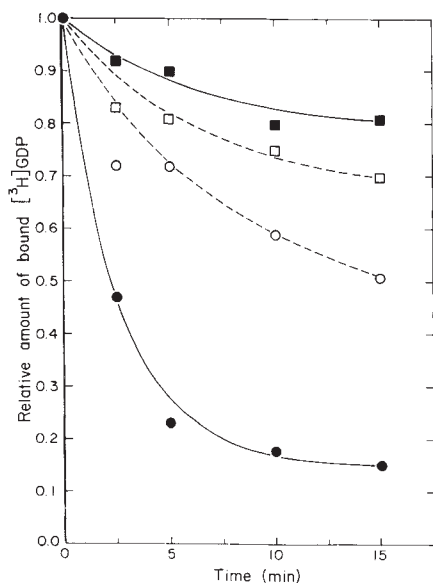


FIG. 2 Effects of the *S. frugiperda* lysates containing p66^{dbl} on [³H]GDP binding to the human platelet CDC42 and *rac1* proteins. Samples (4 µl) from the *S. frugiperda* control lysates (□, ■) or from lysates expressing the p66^{dbl} (○, ●), were added to incubations (75 µl) containing the human platelet CDC42Hs protein (■, ●) or the human platelet *rac1* protein (□, ○) and the time courses for the dissociation of [³H]GDP from these GTP-binding proteins were measured.

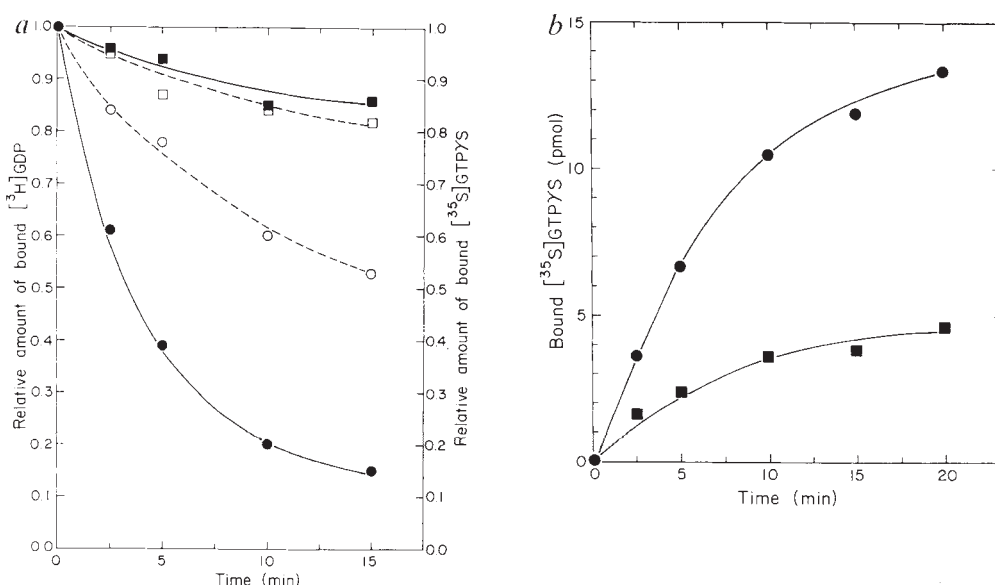
METHODS. *S. frugiperda* cells expressing the p66^{dbl}, and control cells (infected with wild-type baculovirus), were pelleted and the cell pellets were resuspended in 3 ml 20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, plus 0.5 mM PMSF, 5 µg ml⁻¹ leupeptin, 5 µg ml⁻¹ aprotinin and 5 µg ml⁻¹ trans-epoxysuccinyl-L-leucyl-amido(4-guanidino)butane. The resuspended insect cells were homogenized with a Dounce homogenizer and then centrifuged for 15 min at 4 °C in a microfuge. The supernatants represented the cell lysates used in all the experiments. The CDC42Hs protein was solubilized

from human platelet membranes using 1% sodium cholate and purified by a series of chromatography steps which included DEAE-Sephacel, Ultrogel AcA34, hydroxyapatite and Mono-Q, essentially as outlined in Hart *et al.*²¹. Western blot analyses using anti-peptide CDC42Hs antibodies (raised against peptides representing amino acids 167–183 and 167–191) indicated that the CDC42Hs did not bind to the Mono-Q-resin and was present in the flow-through fractions. These fractions were concentrated by applying the protein to a 3 ml hydroxyapatite column, equilibrated in 20 mM Tris-HCl, pH 8.0, and 0.5% CHAPS, and then eluting the CDC42Hs with the same buffer containing 100 mM potassium phosphate, 40% glycerol and 0.5% CHAPS. Preparations of the platelet CDC42Hs used in these experiments were assessed by protein staining following SDS-PAGE to be greater than 90% pure and showed no indication of the presence of the *ras*, *rap1A*, or *rac* proteins, as demonstrated by western blot analyses with antibodies specific for these different GTP-binding proteins. The *rac1* protein used in these experiments was purified from human platelets as described in Polakis *et al.*²². The effects of the *S. frugiperda* cell lysates (control and lysates expressing p66^{dbl}) on the dissociation of [³H]GDP from the purified human platelet CDC42Hs and *rac1* protein were examined by preloading the CDC42Hs (~5 µg) with ~10 µM [³H]GDP (10 Ci mmol⁻¹) for 25 min at room temperature in buffer containing 20 mM Tris-HCl, pH 7.5, 0.1 mM DTT, 0.1 mM EDTA, 3 mM MgCl₂, 50 mM NaCl and 0.1 mM AMP-PNP (final volume 60 µl). Samples from the *S. frugiperda* lysates (typically 4 µl) containing p66^{dbl}, or from control lysates, were added to a second incubation (designated the exchange-assay incubation) which contained 20 mM Tris-HCl, pH 7.5, 0.1 mM DTT, 80 mM NaCl, 5 mM MgCl₂, 0.8 mM AMP-PNP and 1 mM GTP (16 µl final volume). After 15 min at room temperature, 4 µl of the loading incubation containing guanine nucleotide-bound CDC42Hs were added to the exchange-assay incubation. The amount of radiolabelled guanine nucleotide remaining bound to CDC42Hs was then measured at different time points by filtration on nitrocellulose BA85 filters²¹. Experiments describing the effects of p66^{dbl} on [³H]GDP binding to CDC42Hs were done three times with essentially identical results; the data for the effects on [³H]GDP binding to the *rac1* protein are representative of two experiments. In addition, the ability of lysates containing p66^{dbl} to stimulate specifically [³H]GDP dissociation from the platelet CDC42Hs was verified using three different lysates from *S. frugiperda* cells infected with the *dbl* recombinant baculovirus. Three different lysates from control cells infected with the wild-type baculovirus, as well as lysates from uninfected *S. frugiperda* cells, showed no ability to catalyse [³H]GDP dissociation from the platelet CDC42Hs, that is, the results were identical to those obtained when buffer (alone) was added to CDC42Hs.

FIG. 3 Effects of the *S. frugiperda* lysates containing p66^{dbl} on [³H]GDP and [³⁵S]GTPγS binding to the human CDC42Hs protein. *a*, Time course for the dissociation of bound [³H]GDP or [³⁵S]GTPγS from the platelet CDC42Hs protein. Samples (4 μl) from the *S. frugiperda* control lysates (□, ■) or from lysates expressing the p66^{dbl} (○, ●), were added to incubations (20 μl) containing ~5 μg of the human platelet CDC42Hs protein and [³H]GDP (■, ●) or [³⁵S]GTPγS (□, ○). *b*, Time course for the binding of [³⁵S]GTPγS to the platelet CDC42Hs protein. The CDC42Hs protein was preloaded with GDP and then added to incubations (20 μl) containing [³⁵S]GTPγS together with samples from the *S. frugiperda* control lysates (■) or from lysates expressing p66^{dbl} (●).

METHODS. The effects of the lysates from *S. frugiperda* cells expressing p66^{dbl} on the dissociation of the different radiolabelled guanine nucleotides from CDC42Hs were determined as described in the legend for Fig. 2.

The ability of the *S. frugiperda* control lysates, or the lysates containing p66^{dbl}, to catalyse the exchange of GDP for [³⁵S]GTPγS on the human platelet CDC42 protein was determined by first incubating CDC42Hs (5 μg) with 8 μM GDP and 6 mM MgCl₂ for 10 min at room temperature. A sample of the GDP-bound CDC42Hs (~0.3 μg) was then added to an incubation



containing 100 μM AMP-PNP, 10 mM MgCl₂, and 5 μM [³⁵S]GTPγS (~11,000 c.p.m. pmol⁻¹) and 4 μl of either the control lysates from *S. frugiperda* cells or lysates containing the p66^{dbl}. The binding of [³⁵S]GTPγS was measured at the indicated times as described by Hart *et al.*²¹. The results shown in *a* and *b* are representative of two experiments.

of radiolabelled GDP from the platelet CDC42Hs, even after 15 min at room temperature, and yielded results indistinguishable from those obtained with buffer alone. Essentially identical results were obtained when two other sets of *S. frugiperda* cells infected with the recombinant baculovirus (which expresses p66^{dbl}) were compared with cells infected with the wild-type virus. To obtain evidence that it is the p66^{dbl} within the *S. frugiperda* lysates that is stimulating GDP dissociation, we immunoprecipitated the GDP-dissociation activity using antisera raised against amino acids 818–831 of the proto-*dbl* sequence. The results of these studies show that ~80% (83% ± 8%, *n* = 4) of the GDP-dissociation activity measured in the original p66^{dbl} containing lysates could be recovered from precipitated anti-*dbl*-protein A beads, compared with an ~10% recovery of the activity (11% ± 2%, *n* = 2) from nonspecific IgG-protein A beads and no recovery of measurable activity when anti-*dbl*-protein A beads were precipitated from control lysates.

[³H]GDP binding to the recombinant CDC42Hs protein purified from *S. frugiperda* cells was also sensitive to p66^{dbl}. The half-time for the p66^{dbl}-stimulated dissociation of [³H]GDP from the recombinant CDC42Hs occurred within ~2.5 min compared with a half-time of dissociation of greater than 20 min in the presence of control lysates. These results confirmed that the p66^{dbl} catalysed GDP dissociation from the CDC42Hs protein, rather than from a related GTP-binding protein that might have been present in the platelet preparation. In fact, the effects of p66^{dbl} on GDP dissociation seem to be highly specific for the CDC42Hs protein. For example, p66^{dbl} is much less effective in eliciting the dissociation of [³H]GDP from the human platelet *rac1* protein compared with the human platelet CDC42Hs protein (Fig. 2), despite the fact that the amino-acid sequences for the *rac* proteins are 70% identical to that for CDC42Hs⁹. Moreover, under conditions where p66^{dbl} effectively stimulates GDP dissociation from the *S. frugiperda*-recombinant CDC42Hs protein, there is no detectable stimulation of the dissociation of [³H]GDP from the *S. frugiperda*-recombinant c-H-ras or *rap1A* proteins (data not shown).

The effects of p66^{dbl} on the rate of dissociation of a GTP-analogue ([³⁵S]GTPγS) from the CDC42Hs protein were not so marked as the effects on [³H]GDP binding. Under conditions

where p66^{dbl} caused the dissociation of ~85% of the bound [³H]GDP, only ~45% of the bound [³⁵S]GTPγS was dissociated (Fig. 3a). The dissociation of [³⁵P]GTP from CDC42Hs also seems to be only weakly catalysed by the *dbl* protein (data not shown). The reduced capability of p66^{dbl} to elicit the dissociation of radiolabelled GTP-analogues from CDC42Hs suggested that p66^{dbl} would stimulate the binding of radiolabelled GTP (or GTP analogues) to a GDP-bound CDC42Hs species and therefore serve to catalyse guanine nucleotide exchange. The results presented in Fig. 3b illustrate that this is the case. When the CDC42Hs protein was preloaded with GDP and then added to incubations containing either p66^{dbl}, or control lysates, there was a significant stimulation by p66^{dbl} of the binding of [³⁵S]GTPγS to CDC42Hs, which occurred over a time period similar to that for the p66^{dbl}-stimulated dissociation of [³H]GDP.

Our present studies provide the first demonstration of the involvement of an oncogenic protein in the regulation of guanine nucleotide binding to a *ras*-related protein. Given that we find CDC42Hs to be ubiquitously distributed (M.J.H., unpublished results) the apparent tissue-selective expression of proto-*dbl* (brain, adrenal glands and gonads)¹² raises the question of whether other Dbl-related molecules are responsible for regulating guanine nucleotide binding to CDC42Hs in other cells. At present, the role of CDC42 in mammalian cells is not known but, from studies with the yeast CDC42 protein, it seems likely that the mammalian GTP-binding protein is involved in some aspect of cytokinesis and/or cell morphogenesis, perhaps by providing orientational signals to elements of the cytoskeleton. Previous studies have indicated that the *dbl* oncogene and proto-oncogene products are associated with the cytoskeletal matrix²⁰. The *dbl* oncogene lacks detectable homology with other characterized oncogenes which typically exert their effects at the cell surface or within the nucleus. Thus it now seems possible that *dbl* may be representative of a new class of oncogenic proteins that elicit their transforming effects by acting directly on members of the *ras*-related family of GTP-binding proteins, thereby regulating cytoskeletal-associated activities involved in the maintenance of cell polarity and shape and in the induction of cytokinesis. □

Received 18 July; accepted 30 September 1991.

- Hall, A. *Science* **249**, 635-640 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
- Evans, T., Hart, M. J. & Cerione, R. A. *Curr. Opin. Cell Biol.* **3**, 185-191 (1991).
- Evans, T., Brown, M. L., Fraser, E. D. & Northup, J. K. *J. Biol. Chem.* **261**, 7052-7056 (1986).
- Polakis, P. G., Snyderman, R. & Evans, T. *Biochem. biophys. Res. Commun.* **160**, 25-32 (1989).
- Adams, A. E. M., Johnson, D. I., Longnecker, R. M., Sloat, B. F. & Pringle, J. R. *J. Cell Biol.* **111**, 131-142 (1990).
- Johnson, D. I. & Pringle, J. R. *J. Cell Biol.* **111**, 143-152 (1990).
- Munemitsu, S. *et al. Molec. cell. Biol.* **10**, 5977-5982 (1990).
- Shinjo, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9853-9857 (1990).
- Sloat, B. F., Adams, A. & Pringle, J. R. *J. Cell Biol.* **89**, 395-405 (1981).
- Eva, A. & Aaronson, S. A. *Nature* **316**, 273-275 (1985).
- Ron, D., Tronick, S. R., Aaronson, S. A. & Eva, A. *EMBO J.* **7**, 2465-2473 (1988).
- Eva, A., Vecchio, G., Rao, C. D., Tronick, S. R. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2061-2065 (1988).
- Ron, D. *et al. New Biologist*, **3**, 372-379 (1991).
- Wolfman, A. & Macara, I. G. *Science* **248**, 67-69 (1990).
- Huang, Y. K., Kung, H.-F. & Kamata, T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8008-8012 (1990).
- Yamamoto, T. *et al. J. Biol. Chem.* **265**, 16626-16634 (1990).
- Graziani, G., Nebreda, A., Srivastava, S. K., Santos, E. & Eva, A. *Oncogene* (in the press).
- Hart, M. J., Shinjo, K., Hall, A., Evans, T. & Cerione, R. A. *J. Biol. Chem.* (in the press).
- Graziani, G., Ron, D., Eva, A. & Srivastava, S. K. *Oncogene* **4**, 823-829 (1989).
- Hart, M. J., Polakis, P. G., Evans, T. & Cerione, R. A. *J. Biol. Chem.* **265**, 5990-6001 (1990).
- Polakis, P. G. *et al. J. Biol. Chem.* **264**, 16378-16382 (1989).

ACKNOWLEDGEMENTS. We thank C. Westmiller for expert technical assistance and R. Ward for helpful suggestions regarding the manuscript. This work was supported by grants to R.A.C. from the American Cancer Society and the National Institutes of Health.

Role for cyclin A in the dependence of mitosis on completion of DNA replication

Duncan H. Walker & James L. Maller

Howard Hughes Medical Institute, University of Colorado School of Medicine, Denver, Colorado 80262, USA

THE cyclins were first identified by their cell-cycle-dependent synthesis and destruction¹⁻³ and have a key role in the control of mitosis in *Xenopus* embryonic cell cycles⁴⁻⁶. All higher eukaryotes have at least two types of cyclins, the A- and B-type, which can be distinguished by sequence motifs and the timing of their destruction in the cell cycle^{2,7-10}. The degradation of both cyclins is required for exit from mitosis¹¹, but the activation and destruction of cyclin A occur earlier in the cell cycle than with the B-type cyclins⁹⁻¹¹. This suggests that cyclin A has a distinct role in cell-cycle progression. We have used an antisense oligodeoxynucleotide directed against cyclin A to investigate this role. Ablation of cyclin A messenger RNA in cytotstatic factor/metaphase-arrested extracts of *Xenopus* eggs, followed by *in vitro* progression into interphase, resulted in the premature appearance of cyclin B/*cdc2*-associated H1 kinase activity and premature entry into mitosis. Although cyclin A-ablated extracts could initiate DNA synthesis during interphase, S phase was not completed before entry into mitosis. The effects of cyclin A ablation were reversed by the addition of cyclin A mRNA or cyclin A protein to the extracts.

An antisense oligodeoxynucleotide (ODN) was generated against a sequence element of *Xenopus* cyclin A mRNA not present in either *Xenopus* cyclin B1 or B2 mRNAs^{4,10}. When this ODN was added to a wheat-germ lysate translation system (which contains RNase H, ref. 12), it inhibited the efficient translation of cyclin A mRNA, whereas cyclin B2 mRNA translation remained largely unaffected (Fig. 1a). The absence of cyclin A protein in egg extracts after ODN treatment was also confirmed both by blotting with anti-cyclin A antibodies (Fig. 1b) and by measuring the loss of H1 kinase activity of cyclin A immunoprecipitates (Fig. 3). The appearance of cyclin B protein and the level of its associated H1 kinase activity remained unaffected in the same extracts (Figs 1c and 3).

To investigate the role of cyclin A in the cell cycle, cytotstatic factor (CSF)-arrested extracts were incubated with antisense ODN before the addition of 0.4 mM Ca²⁺ to induce exit from

metaphase⁵. Cycles of mitosis and DNA synthesis in these extracts have morphological and biochemical properties similar to *in vivo* cycles although the kinetics of degradation of the *c-mos* proto-oncogene product may be different in intact eggs and extracts^{5,6,10,13-15}. At various times after exit, samples of the extract were taken and assayed for cyclin B/*cdc2*-associated H1 kinase activity (MPF) or scored for NEBD (nuclear envelope breakdown) when appropriate. In the absence of added sperm chromatin in the extracts, MPF/H1 kinase was activated at 100 min both after treatment with cyclin A-antisense ODN and in a control (untreated) cycle (Fig. 2a). Extracts treated with a control 'sense' ODN also had no change in the MPF/H1 kinase profile either with or without added DNA (data not shown).

One of the hallmarks of cell-cycle control is the dependence of mitosis on earlier completion of DNA synthesis^{13,16-18}. This control point is present even in early embryonic cycles. Entry into mitosis in egg extracts can therefore be inhibited by preventing the completion of DNA synthesis in either of two ways. First, extracts can be induced to delay in interphase by the addition of a high concentration of sperm chromatin that cannot be replicated in the normal interphase period¹³. Second, the DNA synthesis inhibitor aphidicolin causes extracts to arrest in interphase in the presence of even small quantities of DNA¹³. When extracts were subjected to either of these conditions, ablation of cyclin A resulted in premature entry into mitosis without completion of DNA synthesis. Figure 2b shows the effect of addition of 15 ng sperm chromatin μl^{-1} (which forms roughly 5,000 pronuclei μl^{-1} , ref. 14) to extracts treated with or without antisense ODN. Chromatin addition caused control extracts to remain in the interphase period for an additional 40 min with a basal level of MPF/H1 kinase activity relative to an extract without added DNA (Fig. 2a). But the MPF/H1 kinase activity in antisense cyclin A ODN-treated extracts increased at the normal time (100 min) independently of the added DNA (Fig. 2b). These changes in H1 kinase activity reflected entry into mitosis as judged by nuclear morphology (Fig. 2d, e).

The ablation of cyclin A protein also overcame an aphidicolin block in extracts containing much smaller amounts of DNA (Fig. 2c). Control and antisense cyclin A ODN-treated extracts were incubated in the presence of 50 $\mu\text{g ml}^{-1}$ aphidicolin and 3 ng sperm chromatin μl^{-1} . These conditions arrested control extracts in interphase with low levels of total H1 kinase activity (Fig. 2c), although the quantitatively minor H1 kinase activity associated with cyclin A in these extracts increased despite the fact that the extracts were arrested in interphase (Fig. 3d). But extracts treated with antisense cyclin A ODN entered mitosis 140 min after CSF release, as determined by an increase in MPF/H1 kinase activity (Fig. 2c) and nuclear envelope breakdown (not shown).

To ensure that the effect of the antisense ODN was specific¹⁹, we added either a bacterially expressed protein A/cyclin A fusion protein or *Xenopus* cyclin A mRNA to antisense cyclin A-treated CSF extracts which had been brought into interphase by calcium addition 10 min earlier. Because the half-life of ODN in *Xenopus* extracts is short (about 10 min²⁰), cyclin A mRNA added at this time is not destroyed, and cyclin A protein was synthesized, indicating that the antisense ODN had been degraded by this time. Readdition of cyclin A mRNA restored the delay in timing of H1 kinase activation to control values (Fig. 2b). Similarly addition of recombinant protein A/cyclin A protein to the aphidicolin-treated extract abolished the activation of H1 kinase caused by cyclin A ODN treatment (Fig. 2c). As a control, preparations from *Escherichia coli* expressing only the protein A vector alone at the same concentration had no effect, and the amount of cyclin A required (10 nM) is well below the concentration of this cyclin needed to induce mitosis in these extracts (100 nM). Similarly, the amount of cyclin produced by added mRNA was slightly lower than that in untreated extracts (data not shown), confirming the sensitivity of this system to added cyclin A.

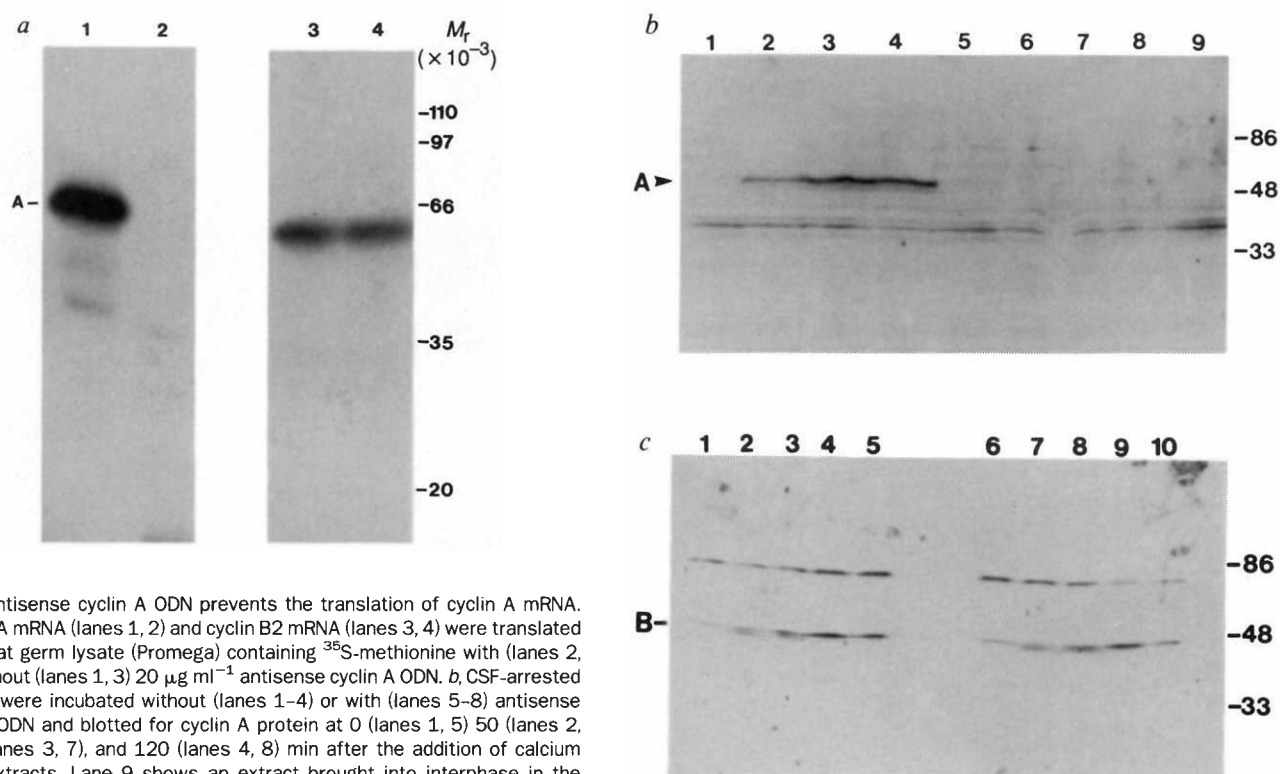


FIG. 1 Antisense cyclin A ODN prevents the translation of cyclin A mRNA. **a**, Cyclin A mRNA (lanes 1, 2) and cyclin B2 mRNA (lanes 3, 4) were translated in a wheat germ lysate (Promega) containing ^{35}S -methionine with (lanes 2, 4) or without (lanes 1, 3) $20\text{ }\mu\text{g ml}^{-1}$ antisense cyclin A ODN. **b**, CSF-arrested extracts were incubated without (lanes 1–4) or with (lanes 5–8) antisense cyclin A ODN and blotted for cyclin A protein at 0 (lanes 1, 5) 50 (lanes 2, 6), 90 (lanes 3, 7), and 120 (lanes 4, 8) min after the addition of calcium to the extracts. Lane 9 shows an extract brought into interphase in the presence of $25\text{ }\mu\text{g ml}^{-1}$ cycloheximide and $100\text{ }\mu\text{M}$ emetine. **c**, CSF-arrested extracts were incubated without (lanes 1–5) or with (lanes 6–10) antisense cyclin A ODN and blotted for cyclin B2 at 0 (lanes 1, 6), 50 (lanes 2, 7), 90 (lanes 3, 8), 120 (lanes 4, 9), and 150 (lanes 5, 10) min after the addition of calcium to cause exit from metaphase.

METHODS. The antisense ODN against cyclin A (3'-CAAAGTGGGACAATC-5') was directed against the nucleotides encoding the sequence DCPSL in cyclin A protein (single-letter amino-acid code). A sense ODN (3'-GATTGTCCAG-TTTG-5') was used as a control. mRNA was transcribed from pGEM constructs using a Promega kit, and wheat germ translations of the mRNA were done with another Promega kit using an mRNA concentration of $20\text{ }\mu\text{g ml}^{-1}$. Samples of each translation assay were run on 10% SDS polyacrylamide

gels²⁸, after which the gels were stained, washed for 30 min in 1M sodium salicylate, dried and exposed to X-ray film. The arrows marked A and B indicate the positions of cyclins A and B2, respectively. CSF (metaphase-arrested) extracts were prepared as described previously^{6,11} except that leupeptin was omitted from all buffers. Aliquots of the extracts were incubated with $20\text{ }\mu\text{g ml}^{-1}$ antisense cyclin A ODN for 45 min at $4\text{ }^{\circ}\text{C}$ with rotation. Sperm chromatin (3 ng ml^{-1}) was added with 0.4 mM CaCl_2 to induce exit from CSF arrest. For blotting, aliquots of extracts at the indicated times were subjected to 10% SDS PAGE, transferred to nitrocellulose and blotted with an anti-cyclin A or anti-cyclin B2 antibody raised in sheep, using the ECL detection method (Amersham).

Cyclin A/*cdc2* was previously thought to be a mitotic cyclin in early *Xenopus* embryonic cycles, although cyclin A, uncoupled to *cdc2*, has been suggested to have a role in S phase¹⁵. To determine if the effects of ablation of cyclin A may be due to an active cyclin A/*cdc2* kinase complex, we assayed H1 kinase activity associated with cyclin A/*cdc2* and cyclin B/*cdc2* in DNA-arrested extracts. Cyclin A/*cdc2* H1 kinase activity accumulated throughout the time of incubation of a control extract (Fig. 3a), whereas the H1 kinase activity associated with cyclin B/*cdc2* did not increase until after 140 min of incubation. These kinetics of activation are similar to those reported by others¹⁰.

Earlier treatment of the extract with antisense cyclin A ODN removed cyclin A/*cdc2* H1 kinase activity and led to the premature appearance of cyclin B/*cdc2* H1 kinase activity at 100 min (Fig. 3b). The addition of cyclin A mRNA to ablated extracts restored both cyclin A/*cdc2* H1 kinase and the delayed kinetics for activation of cyclin B/*cdc2* (Fig. 3c). Figure 3d shows a different extract treated with aphidicolin and low levels of nuclei. Cyclin A/*cdc2* kinase accumulated continuously in the presence of aphidicolin whereas cyclin B never became active. But in the same extracts with cyclin A ablated (Fig. 3e), cyclin B/*cdc2* was activated.

Taken together, these data indicate that the effect of the antisense cyclin A ODN is the result of removal of cyclin A-mediated inhibition of cyclin B/*cdc2* complex (pre-MPF) activation. This provides a potential explanation for the absence of cyclin A in MPF purified from various sources^{21,22}, and

indicates that the activation of the A- and B-type cyclins has very different consequences in cell-cycle progression: activation of cyclin A/*cdc2* in interphase maintains this stage of the cell cycle until DNA synthesis is complete, whereas the activation of cyclin B/*cdc2* causes entry into mitosis⁶.

The cell-cycle restriction point controlling initiation of DNA synthesis has been attributed to *cdc2* or a *cdc2*-like protein²³. Human cyclin A associates with both *cdc2* and a *cdc2*-like protein, p33 (ref. 9), although only genuine *cdc2* has been found associated with cyclin A in *Xenopus*^{10,24}. A *cdc2*-related kinase known as *Eg1* is present in *Xenopus* but is not associated with either cyclin A or cyclin B^{24,25}. As cyclin A/*cdc2* H1 kinase activity normally appears during DNA synthesis before that of cyclin B (refs 9, 10), we investigated the effect of cyclin A ablation on DNA synthesis. The incorporation of [α - ^{32}P]dATP into sperm DNA was similar in control, cycloheximide and 'sense' ODN-treated extracts (Fig. 4). DNA synthesis can occur in the absence of any cyclins¹⁴ and thus suggests that cyclin A has no essential role in the process of DNA synthesis. But the incorporation of [α - ^{32}P]dATP in antisense cyclin A ODN-treated extracts was less than 40% of that seen in controls in several experiments (Fig. 4), although some DNA synthesis was always observed. Kinetic studies demonstrated DNA synthesis is initiated normally in antisense-treated extracts, but terminates prematurely because of entry into mitosis (data not shown). The addition of cyclin A protein to antisense-treated extracts restored DNA synthesis to near control levels (Fig. 4, bar D) by lengthening the interphase period of the cell cycle.

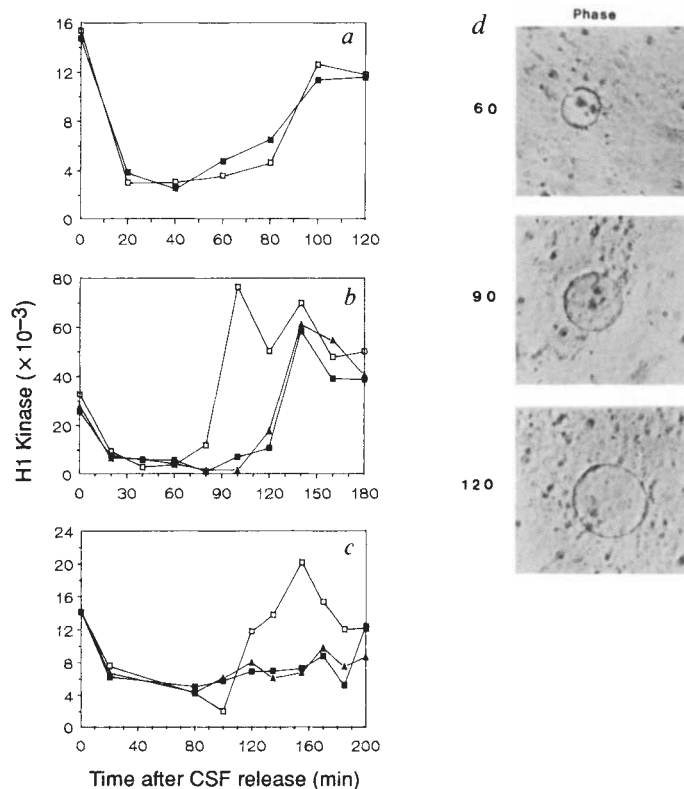


FIG. 2 Cyclin A ablation causes premature mitotic entry in extracts. Untreated (■) and antisense treated (□) CSF-arrested extracts were incubated without added DNA (a) with 15 ng μl^{-1} DNA (b), or with 3 ng μl^{-1} DNA and 50 $\mu\text{g ml}^{-1}$ aphidicolin (c). b, Effect of readdition of cyclin A mRNA 10 min after CSF release (▲). c, Effect of addition of cyclin A protein 20 min after CSF release (▲). At various times after calcium-induced release from CSF arrest, aliquots were removed and assayed for H1 kinase activity. d, e, Phase-contrast (Phase) and fluorescence (Fluor) microscopy of nuclei done under conditions as in b. d, Nuclei in a control extract. e, Nuclei in an

antisense-treated extract at 60, 90 and 120 min after exit from metaphase. METHODS. CSF-arrested extracts were prepared and, where indicated, aphidicolin (50 $\mu\text{g ml}^{-1}$) was added 10 min before calcium addition. The control and antisense-treated extracts were incubated with sperm chromatin and 0.4 mM CaCl_2 to induce exit from CSF arrest. Samples were removed at various times and assayed for either H1 kinase activity or for nuclear morphology both by phase-contrast and by fluorescence microscopy as described²¹. *E. coli* carrying a bovine protein A/cyclin A fusion protein vector or the protein A vector¹⁵ alone were a gift of T. Hunt. After induction with IPTG, the cells were lysed and centrifuged at 20,000g for 30 min. The concentration of the protein A/cyclin A construct was determined by comparison with standards run in adjacent lanes of SDS gels as described. The concentration of protein used in c (10 nM), activated H1 kinase activity in cycloheximide-treated interphase extracts to a level roughly 5% of that seen in mitosis in untreated extracts. Controls with the protein A vector alone at the same concentration had no effect.

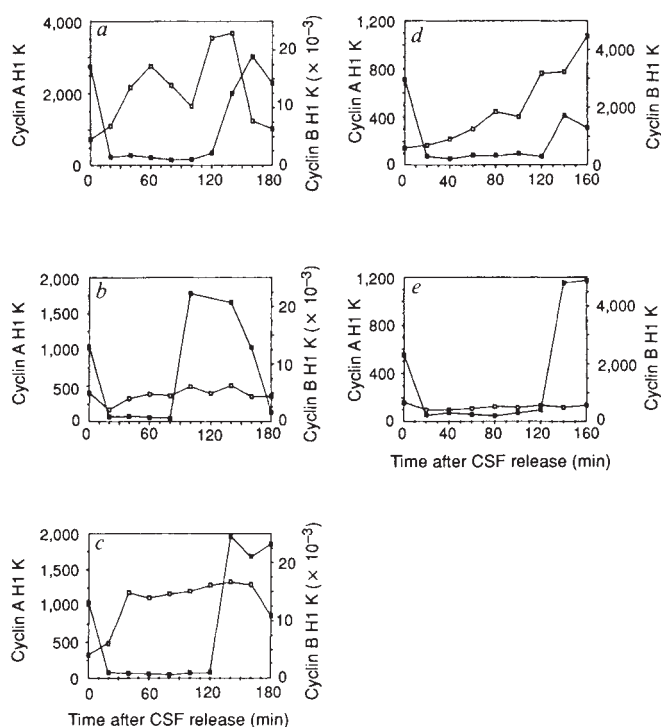


FIG. 3 Addition of cyclin A mRNA to ablated extracts restores feedback control of mitosis on DNA synthesis. CSF-arrested extracts were treated without (a) or with (b) antisense cyclin A ODN and incubated with 15 ng μl^{-1} sperm chromatin and 0.4 mM CaCl_2 . Cyclin A mRNA at a final concentration of 20 $\mu\text{g ml}^{-1}$ was added to a second antisense-treated sample 10 min after the addition of CaCl_2 (c). In another set of experiments with different extracts, samples from control (d) and antisense cyclin ODN-treated (e) extracts were incubated with 3 ng μl^{-1} sperm chromatin and 50 $\mu\text{g ml}^{-1}$ aphidicolin before addition of CaCl_2 to induce exit from CSF arrest. At various times samples were removed and cyclin A (□) and cyclin B (■) immunoprecipitates were prepared and assayed for H1 kinase activity. The level of cyclin B/*cdc2* H1 kinase activity at 140 min in d was less than 20% of the mitotic level and did not cause any changes in nuclear morphology. Because we used different extracts with different [γ -³²P]ATP-specific activities in a-c and d, e, no direct comparison of H1 kinase activity is possible.

METHODS. Antibodies were generated in sheep against bacterially expressed cyclins A, B1 and B2. Immunoprecipitates were done as described²⁸ except that protein G-Sepharose beads were used instead of protein A-Sepharose beads. Cyclin B immunoprecipitates used a mixture of antisera generated against either cyclin B1 or B2. The immunoprecipitates were assayed for H1 kinase and electrophoresed on 12.5% SDS polyacrylamide gels as described²⁸. The gels were stained, dried, and the H1 bands were excised and counted.

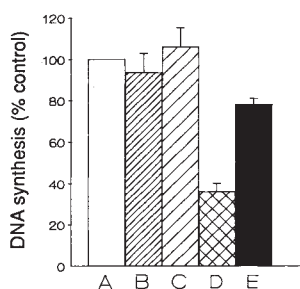


FIG. 4 Premature termination of DNA synthesis after ablation of cyclin A. CSF-arrested extracts were treated with no additions (A), with 25 $\mu\text{g ml}^{-1}$ cycloheximide (B), with sense cyclin A ODN (C), antisense cyclin A ODN (D) or antisense cyclin A ODN to which 10 nM bacterially expressed protein A/cyclin A fusion protein was added 20 min after the addition of calcium (E) and assayed for DNA synthesis as described¹⁴. Briefly, sperm chromatin at 30 $\text{ng } \mu\text{l}^{-1}$ and [α -³²P]dATP at 20 $\mu\text{Ci ml}^{-1}$ were added to the extracts when CSF release was induced by calcium addition. Assays were terminated after 3 h by the addition of 1 vol. 50 mM Tris pH 7.5, 5 mM EGTA, 0.1% SDS, 0.5 mg ml^{-1} proteinase K. Samples were incubated for 1 h at 37 °C and then extracted with phenol-chloroform, TCA precipitated spotted onto glass fibre filters, washed and counted for incorporation of ³²P into DNA. All experiments were normalized to a control in which no DNA has been added. The results of three experiments are expressed as the average $\pm$ s.e.m.

There is a link between completion of DNA synthesis and the activation of the *cdc25* phosphatase^{16–18}. We suggest the ablation of cyclin A may prematurely activate *cdc25* which would activate cyclin B/*cdc2* by dephosphorylation of Tyr 15. Although the ablation of cyclin A overcomes an 'incomplete DNA synthesis block' to MPF activation, no effect of cyclin A on cyclin B/*cdc2* activation is evident in the absence of DNA (compare Fig. 2a with b). This is probably because of a requirement for the association of cyclin A with the nucleus. Cyclin A associates with the nucleus in somatic cells²⁷ and we find that most endogenous cyclin A rapidly associates with the nuclear fraction in cycling extracts (data not shown). Note that *cdc25* also seems to act in the nucleus (refs 17, 18 and data not shown). The data in Fig. 2 also implies that the activation of cyclin B/*cdc2* is dependent on other controls, such as *wee1*, *cdc25* or *RCC1* (ref. 26) as well as cyclin A/*cdc2*.

These results suggest that cyclin A has a function during S phase in regulating the activation of cyclin B/*cdc2*. Cyclin A would be required for the normal transition from S phase to mitosis. But note that other potential roles for cyclin A have been proposed^{8,11,15}. The persistent expression of active cyclin A/*cdc2* H1 kinase throughout S phase and early M phase in early cell cycles (Fig. 3 and refs 9, 10) suggests it has many roles in ensuring proper cell-cycle progression. Therefore the ablation of cyclin A may manifest itself in different ways in different systems, depending on the presence or absence of other control points in those systems. It was previously unclear why there are two classes of G₂ cyclins and why cyclin A is both activated and degraded before cyclin B. These results present direct evidence for a role of cyclin A in cell-cycle progression distinct from that of cyclin B and implicate it in the feedback control between mitosis and DNA synthesis. □

14. Blow, J. J. & Laskey, R. A. *Cell* **47**, 577–587 (1986).
15. Bandara, L. R., Adamczewski, J. P., Hunt, T. & La Thangue, N. B. *Nature* **352**, 249–251 (1991).
16. Gould, K. & Nurse, P. *Nature* **342**, 39–45 (1989).
17. Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
18. Kumagai, A. & Dunphy, W. G. *Cell* **64**, 903–914 (1991).
19. Colman, A. J. *Cell Sci.* **97**, 399–409 (1990).
20. Dagle, J. M., Walder, J. A. & Weeks, D. L. *Nucleic Acids Res.* **18**, 4751–4757 (1990).
21. Lohka, M. J., Hayes, M. K. & Maller, J. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3009–3013 (1988).
22. Labbé, J.-C. *et al.* *EMBO J.* **8**, 3053–3058 (1990).
23. Blow, J. J. & Nurse, P. *Cell* **62**, 855–862 (1990).
24. Gabrielli, B. G., Roy, L. M., Gautier, J., Philippe, M. & Maller, J. L. *J. biol. Chem.* (in the press).
25. Solomon, M. J., Glotzer, M., Lee, T. H. & Kirschner, M. W. *Cell* **63**, 1013–1024 (1990).
26. Nishitani, H. *et al.* *EMBO J.* **10**, 1555–1564 (1991).
27. Pines, J. & Hunter, T. *J. Cell Biol.* **115**, 1–17 (1991).
28. Gautier, J. *et al.* *Cell* **60**, 487–494 (1990).

ACKNOWLEDGEMENTS. We thank T. Hunt (ICRF Clare Hall Laboratories, UK) for *Xenopus* cyclin A, B1 and B2 cDNAs and the bovine protein A cyclin A fusion protein and for suggesting the choice of antisense cyclin A ODN. T. Izumi for help with *in vitro* synthesis of mRNA and M. J. Lohka and E. Erikson for critical reading of the manuscript. This work was supported by the NIH and the Lucille P. Markey Charitable Trust. J.L.M. is an Investigator of the Howard Hughes Medical Institute.

DNA-binding properties of the product of the testis-determining gene and a related protein

Nargis Nasrin, Colleen Buggs, Xiang Fu Kong*, James Carnazza, Mark Goebel* & Maria Alexander-Bridges†

Howard Hughes Medical Institute, Diabetes Unit and Medical Services, Massachusetts General Hospital, Harvard Medical School, 50 Blossom Street, Boston, Massachusetts 02114, USA

THE upstream region of the human glyceraldehyde-3-phosphate dehydrogenase gene contains an insulin-response element (IRE-A) responsible for insulin-dependent transcription of the gene (ref. 1). The open reading frame of a rat complementary DNA encoding a protein (IRE-ABP) that binds to this sequence contains an HMG box motif² that is 67% identical to the mouse candidate gene for the testis-determining factor SRY, and 98% identical to the mouse SRY-like gene, *a4* (refs 3, 4). Here we report that IRE-ABP and SRY bind to IRE-A DNA with comparable specificity in a DNase-I footprinting assay. Two females with sex reversal were found to have a single amino-acid substitution in the HMG box domain of SRY at position 3 and 7, respectively^{5,6}. SRY derivatives containing corresponding mutations do not make contact with IRE-A DNA. These results are direct evidence that mouse SRY-like proteins are sequence-specific DNA-binding proteins and identify two amino acids critical to this interaction. Moreover, IRE-A is a candidate SRY-response element.

We have shown that insulin stimulates transcription of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene in 3T3 adipocytes and we have mapped the response element IRE-A to nucleotides –488 to –435 in the 5'-flanking region of this gene¹. The ability of an insulin-responsive protein, IRP-A, to bind IRE-A DNA correlates with the ability of this element to make gene transcription insulin-responsive¹. We used the wild-type IRE-A oligonucleotide to screen^{7,8} a λ gt11 cDNA library prepared from rat adipocytes to isolate a clone encoding an IRE-A DNA-binding protein, IRE-ABP. The tissue distribution of this protein and its chronic regulation by insulin correlate with the pattern of insulin-regulated GAPDH gene expression (our unpublished results), but it is not clear whether IRE-ABP is identical to IRP-A or is a different protein with similar binding specificity.

Rat IRE-ABP and mouse SRY (refs 3, 4) are 68% identical

Received 5 July; accepted 11 October 1991.

1. Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D. & Hunt, T. *Cell* **33**, 389–396 (1983).
2. Swenson, K. I., Farrell, K. M. & Ruderman, J. V. *Cell* **47**, 861–870 (1986).
3. Standart, N., Minshull, J., Pines, J. & Hunt, T. *Dev. Biol.* **124**, 248–258 (1987).
4. Minshull, J., Blow, J. J. & Hunt, T. *Cell* **56**, 947–956 (1989).
5. Murray, A. W. & Kirschner, M. W. *Nature* **339**, 275–280 (1989).
6. Murray, A. W., Solomon, M. J. & Kirschner, M. W. *Nature* **339**, 280–286 (1989).
7. Pines, J. & Hunt, T. *EMBO J.* **6**, 2987–2995 (1990).
8. Lehner, C. F. & O'Farrell, P. H. *Cell* **56**, 957–968 (1989).
9. Pines, J. & Hunter, T. *Nature* **346**, 760–763 (1990).
10. Minshull, J., Golsteyn, R., Hill, C. S. & Hunt, T. *EMBO J.* **9**, 2865–2875 (1990).
11. Roy, L. M. *et al.* *J. Cell Biol.* **113**, 507–514 (1991).
12. Minshull, J. & Hunt, T. *Nucleic Acids Res.* **16**, 6343–6451 (1987).
13. Dasso, M. & Newport, J. W. *Cell* **61**, 811–823 (1990).

* Present addresses: Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA (X.F.K.); Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 638 Barnhill Drive, Indianapolis, Indiana 46202, USA (M.G.).

† To whom all correspondence should be addressed.

FIG. 1 Expression of IRE-ABP and SRY HMG box domains as glutathione-S-transferase fusion genes. The HMG domains of mouse SRY (ref. 3), mouse a4 (ref. 3) and rat IRE-ABP are compared (single-letter amino-acid code). The position of mutations associated with sex reversal in patients J.N. and A.A. are indicated⁵.

		1	3	7		40
Mouse SRY	LESME	GHV	KRPMNAFMVWSR	GERHKL	AOGNPSMONT	ESKQLGCR
Mouse a4	CKTPS	GHI	KRPMNAFMVWSQ	IERRKI	MEQSPDMHNAEI	SKRLGKR
Rat IRE-ABP	CKTPS	GHI	KRPMNAFMVWSQ	IERRKI	MEQSPDMHNAEI	SKRLGKR
		L	I			
Patients		(J.N.)	(A.A.)			
		41			79	
Mouse SRY	WKSL	TEAEKRPFFQEAQRLKTLHREKYPNYKYQPHRRAK				VSQRS
Mouse a4	WKLL	KDSQDKIPPIQEAERLRLLKHMADYPDYKYRPRKKVK				SG...
Rat IRE-ABP	WKLL	KDSQDKIPPIQEAQGLRLKHMADYPDYKYRPRKKVK				SGNTGA

Shaded areas denote identical amino acids. The two amino acids that differ between mouse a4 (E, R) and IRE-ABP (D, G) are underlined.

METHODS. Polymerase chain reactions were performed with one of four 5'-primers: wild-type IRE-ABP (I, M) primer (5'-CCCGGGTCCATGGGCCACATC-AAGCGGCCCATGAACGC-3'); primers that encode mutant proteins IRE-ABP (I→V) (5'-CCCGGGTCCATGGGCCACGCTCAAGCGGCCCATGAACGC-3'), IRE-ABP (I→L) (5'-CCCGGGTCCATGGGCCACCTCAAGCGGCCCATGAACGC-3'), or IRE-ABP (M→I) (5'-CCCGGGTCCATGGGCCACATCAAGCGGCCCATGAACGC-3'); and a single primer designed to include amino acids at the 3' end of the IRE-ABP HMG box (5'-GGATCCAAGCTTCTACCTTCTTCGCGCGCGG-3'). Mouse SRY (V, M) was primed with two oligonucleotides: the forward expression primer was (5'-CCCGGGTCCATGGGCCATGCTCAAGCGGCCCATGAATGCATTT-3'), and the reverse expression primer was (5'-GGATCCAAGCTTCTTCATTTAGCCCTCCGATGAGG-3'). The SRY primers are derived from the SRY sequence^{3,4} and include coding sequence for amino acids 1 to 79. In four independent clones isolated from two mice, we documented a single base change from the published nucleotide sequence of SRY that results in an Ile→Thr

substitution at position 61. This amino acid is located in a region of the HMG box domain that is not conserved between family members. GST/SRY with mutations at position 3 were primed with the SRY (V, M) primer but for GST/SRY (V→I), GTC was changed to ATC, and for GST/SRY (V→L), GTC was changed to CTG; for GST/SRY (M→I), ATG at position 7 was changed to ATA. The polymerase chain reactions were performed using 10 µg mouse genomic DNA to prime the synthesis of SRY, or 2 ng purified DNA isolated from a λgt11 clone that encodes IRE-ABP, in 200 µM each dNTP, 0.5 µM each primer, 1.5 mM MgCl₂, 10 mM Tris (pH 8.3), 50 mM KCl, 8% dimethylsulphoxide, and 0.25 U of *Taq* polymerase in a volume of 100 µl. Reactions were cycled for 0.5 min at 94 °C, 1 min ramp to 55 °C, 0.5 min at 55 °C, 0.5 min ramp to 72 °C, 0.5 min at 72 °C, 1 min ramp to 94 °C, for 25 cycles. Two rounds of PCR were performed on genomic DNA. Products were extracted and digested with *NcoI* and *HindIII* and cloned into PGEX-GT (gift from J. Dixon). The sequence of the cloned fragments was documented to encode amino acids 1–79 as displayed.

over the 79-amino-acid HMG box region (Fig. 1). Four unidentified autosomal genes, *a1–a4*, were isolated by screening a whole mouse embryo cDNA library with SRY-related sequences³. The most divergent, *a4*, was 77% identical to mouse SRY. Rat IRE-ABP is virtually identical to the mouse a4 sequence, except for one conservative and one non-conservative change (Fig. 1).

Because of the similarity between IRE-ABP and SRY in the HMG box region, we investigated whether this segment of SRY conferred sequence-specific binding to the IRE-A motif; we expressed amino acids 1–79 of the HMG box domain of SRY and IRE-ABP as part of a fusion protein with glutathione-S-transferase (GST) (Figs 1 and 2).

Affinity-purified GST/IRE-ABP and GST/SRY proteins bind with similar sequence specificity to the IRE-A motif in a DNase I footprint assay (Fig. 2, lanes 3 and 4). An additional hypersensitive band is evident with active SRY derivatives (Fig. 2, lane 4). Thus, the a4 and SRY-like HMG box domains specifically contact IRE-A DNA. The protection pattern was identical when the HMG box motifs were released from GST with thrombin (data not shown).

The SRY locus is thought to encode a DNA-binding protein that initiates testis formation in the developing embryo^{5,9}. The DNA-binding properties, the potential dimerization domains, and the mechanism by which the gene product regulates transcription are undefined. Three patients with an XY genotype who failed to differentiate to the male phenotype had mutations in the HMG box domain of SRY (refs 5, 6), the putative DNA-binding domain of SRY. Patient J.N. and her father shared a Val→Leu substitution at position 3 (Fig. 1). Thus, it was not possible to establish whether this mutation is completely unrelated to the sex-reversed phenotype or contributes to it. Patient A.A. exhibited a *de novo* mutation causing a Met→Ile substitution at position 7 (Fig. 1); we infer that this mutation is responsible for the sex-reversed phenotype. The third mutation was a four-base deletion that causes a frame shift at position 65 (ref. 6).

Because the response element and target genes of SRY are not known, the functional effect of these mutations on SRY could not be evaluated directly⁵. To determine whether amino acids at positions 3 and 7 are critical for binding of the HMG box family to DNA, we introduced substitutions analogous to

those associated with sex reversal⁵ into the HMG box domains of IRE-ABP and SRY to investigate their effect on affinity for the IRE-A motif. Wild-type IRE-ABP (Ile, Met) specifically contacts IRE-A DNA at protein concentrations ranging from 0.05 to 5 µg per reaction (Fig. 3, lanes 4–7). In the setting of the IRE-ABP HMG box domain, 5 µg of the sex-reversal derivatives IRE-ABP (Ile→Leu) and IRE-ABP (Met→Ile) failed to protect IRE-A DNA from DNase I digestion (Fig. 3, lanes 8 and 9). These mutations therefore cause a 30- to 100-fold decrease in apparent affinity for IRE-A DNA. Similar results were seen when mutations were introduced into the SRY HMG box domain (Fig. 2, lanes 5 and 6). We conclude that introduction of the sex-reversal mutations at positions 3 and 7 in the (SRY-like) HMG box domain significantly reduces binding of this family of proteins to DNA. The observation that a Val→Leu substitution at position 3, found in J.N. and her father, could be responsible for the sex-reversed phenotype supports the suggestion⁵ that J.N.'s father does not express the sex-reversed phenotype because he is mosaic or because the mutation is contributory but not sufficient.

The ability of SRY polypeptides to retard the migration of an IRE-A oligonucleotide is poor compared with IRE-ABP (our unpublished observations). SRY and the SRY-like proteins a1–a3 have a valine at position 3 of the HMG box domain, whereas a4 and IRE-ABP have isoleucine instead (ref. 3). A derivative of SRY with isoleucine substituted for valine at position 3, GST/SRY (V→I), shifts IRE-A DNA to the same extent as does IRE-ABP (our unpublished observations) and protects IRE-A DNA at concentrations comparable to those used for IRE-ABP in the DNase footprint assay (Fig. 3, lanes 10 and 11). These findings imply that SRY and a1–a3 (ref. 3) may have reduced affinity for IRE-A compared with a4 and IRE-ABP.

TABLE 1 Comparison of DNA-binding motifs for members of the HMG-domain family of transcriptional regulators

TCF-1 (ref. 10)	3'-G CTTTG T T-5'
TCF-1α, LEF-1 (refs 11, 12)	5'-(C/T)CTTTG(A/T)-3'
IRE-ABP, SRY	5'-C CTTTG A A-3'

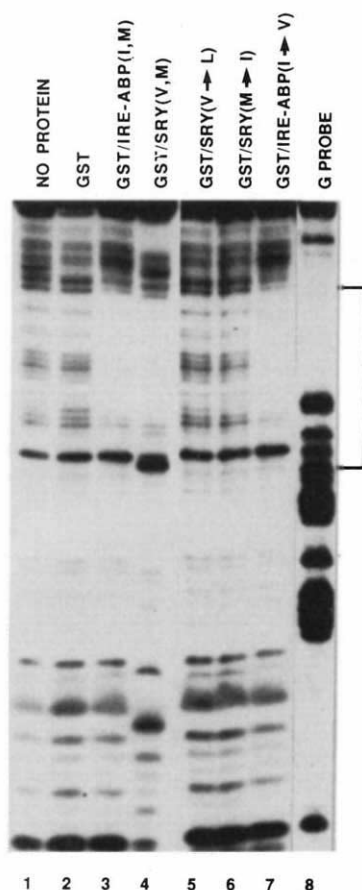


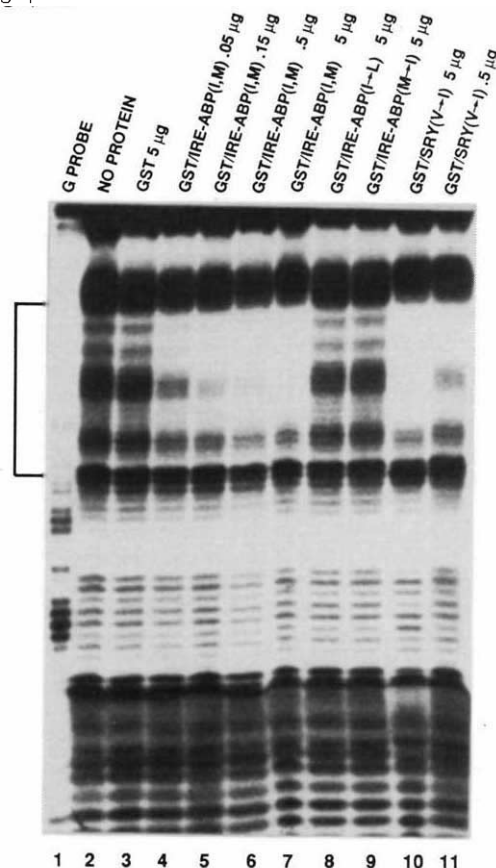
FIG. 2 DNase footprint analysis of GST/SRY derivatives and GST/IRE-ABP. The analysis was performed using wild-type IRE-A DNA (nucleotides -488 to -408) labelled on the antisense strand and (1) purified wild-type GST/SRY (V_3M_7) and GST/IRE-ABP (I_3M_7) proteins; (2) derivatives constructed to mimic

mutations found at positions 3 and 7 of the SRY HMG domain in sex-reversed patients GST/SRY ($V \rightarrow L$) and GST/SRY ($M \rightarrow I$), respectively; and (3) a derivative of GST/IRE-ABP with valine at position 3 to mimic SRY, GST/IRE-ABP ($I \rightarrow V$). The nucleotides protected from DNase I are bracketed (-433(CCCCT-TTCTTTCTTCAAGGCTGAGAG) -460; core consensus sequence in bold; see Table 1). Lane 1, no protein; lane 2, GST; lane 3, wild-type GST/IRE-ABP (I, M); lane 4, wild-type GST/SRY (V, M); lane 5, GST/SRY, with a sex reversal mutation at position 3 ($V \rightarrow L$); lane 6, GST/SRY, with a sex-reversal mutation at position 7 ($M \rightarrow I$); lane 7, GST/IRE-ABP with a switch to valine at position 3 to mimic SRY, IRE-ABP ($I \rightarrow V$); lane 8, G probe, labelled on the antisense strand.

METHODS. Overnight cultures of bacteria transformed with a glutathione expression vector, PGEX-GT, or glutathione fusion genes encoding GST/IRE-ABP or GST/SRY derivatives, were diluted 1:10 in L broth supplemented with $50 \mu\text{g ml}^{-1}$ ampicillin and grown for 1 h at 37°C . After induction with 1 mM isopropyl- β -D-thiogalactoside (IPTG) for 3 h, cells were pelleted by centrifugation at $5,000g$ and extracted. Pelleted cells from a 100-ml culture were lysed by sonication for 15 s in 2 ml buffer A containing 150 mM NaCl, 16 mM NaH_2PO_4 , 4 mM Na_2HPO_4 , 1% Triton X-100, 1.0 mM PMSF, 20 μM Trasylol, 20 μM pepstatin and 20 μM leupeptin, pH 7.4. Cellular debris was removed by centrifugation at $2,000g$ for 5 min. Glutathione Sepharose-4B beads (Pharmacia) were washed with 5–10 bed volumes of buffer A. Bacterial lysates were incubated with 0.125 ml of pelleted beads and mixed gently for 10 min at 4°C . Protein-bound beads were collected and washed twice with 5 bed volumes of buffer A without Triton X-100. Bound proteins were eluted with 5×0.3 ml of elution buffer containing 5 mM glutathione in 50 mM Tris-HCl, pH 8.0, 1 mM DTT, 10% glycerol, 0.5 mM PMSF, 20 μM Trasylol, 20 μM pepstatin, and 20 μM leupeptin. Protein concentration was determined by BioRad microassay, and each preparation was shown to be $>90\%$ homogenous and free of proteolysis by electrophoresis on a 12.5% SDS-polyacrylamide gel. Purified GST fusion proteins isolated in parallel (12 μg) were incubated with an end-labelled fragment of IRE-A DNA labelled on the antisense strand that contained nucleotides (-488 to -408) of the GAPDH promoter in a 50 μl reaction volume containing 20% glycerol, 25 mM HEPES, pH 7.9, 6.25 mM MgCl_2 , 150 mM NaCl and 0.5 mM DTT. After incubation for 60 min on ice, DNase I diluted in 25 mM CaCl_2 was added and the incubation continued at room temperature for 1 min. Samples containing protein were digested with 3 ng DNase I in 3 mM CaCl_2 and samples without protein were digested with 1 ng DNase I. The reaction was terminated by the addition of phenol, re-extracted with phenol/chloroform, precipitated with ethanol and resuspended. 10,000 c.p.m. were loaded onto an 8% sequencing gel and autoradiographed.

FIG. 3 Characterization of GST/IRE-ABP and GST/SRY derivatives. ^{32}P -labelled IRE-A DNA was incubated with concentrations of purified GST/IRE-ABP that varied from 50 ng to 5 μg and the DNase I assay performed as described in the legend to Fig. 2. Lane 1, G probe labelled on the antisense strand; lane 2, no protein; lane 3, 5 μg GST protein; lane 4–7, 0.05 μg , 0.15 μg , 0.5 μg and 5 μg wild-type GST/IRE-ABP (I, M); lane 8, 5 μg GST/IRE-ABP with a sex-reversal mutation at position 3 ($I \rightarrow L$); lane 9, 5 μg GST/IRE-ABP, with a sex-reversal mutation at position 7 ($M \rightarrow I$); lane 10, 5 μg of GST/SRY, with a switch of valine to isoleucine at position 3 to mimic IRE-ABP, GST/SRY ($V \rightarrow I$); lane 11, 0.5 μg GST/SRY, ($V \rightarrow I$).

METHODS. Proteins were prepared as described in the legend to Fig. 2, except that IPTG-induced cells were pelleted by centrifugation at $5,000g$ and extracted with guanidine-HCl¹³ with the following modifications. Pelleted cells were lysed by sonication for 15 s in 6 M guanidine-HCl dissolved in buffer B which contained 50 mM NaCl, 20 mM Tris-HCl, (pH 8.0), 1 mM EDTA, 1 mM DTT, 10% glycerol, 1.0 mM PMSF, 20 μM Trasylol, 20 μM pepstatin and 20 μM leupeptin. Cellular debris was removed by centrifugation at $5,000g$. Supernatants were dialysed in 1.5 M guanidine-HCl hydrochloride in buffer B for at least 2 h, then in buffer B alone overnight. Proteins were purified to 90% homogeneity as described in Fig. 2 legend and were free of degradation.



We conclude that the a4-like gene product, IRE-ABP, is a member of the HMG box family which is highly specific for the IRE-A motif, and that SRY and a1-a3 are sequence-specific DNA binding proteins that interact with a similar but probably distinct range of sequences.

Two T lymphocyte-specific DNA-binding proteins containing an HMG motif have been characterized¹⁰⁻¹². These factors regulate transcription through motifs that are highly conserved with the region of IRE-A DNA that is protected by SRY and IRE-ABP, CCTTTGAA. TCF-1, a factor that activates the CD3 ϵ promoter, makes contact with the sequence GCTTTGTT (ref. 10). TCF1- α (ref. 11), also known as LEF-1 (ref. 12), can activate transcription of the T-cell receptor (TCR) α -chain enhancer. The contact points of LEF-1 with the TCR- α enhancer¹² are contained within a motif, CCTTTGAA, which is identical to the motif protected in IRE-A (Table 1). The core motif 5'-PyCTTTG(A/T)-3' is present in many genes, including those encoding TCR- α , TCR- δ , CD3 (γ and δ), CD4, and p56^{lck}, all of which are bound with high affinity by TCF-1 α (refs 11, 13).

Mutations in this motif that impair binding affinity also inhibit transcriptional activation through the TCR- α enhancer^{11,12}. These observations support the hypothesis that the HMG-box family of proteins modulates transcription through sequences that contain the core motif 5'-Py-CTTTG(A/T)-3', and may help identify targets of the SRY-like family of transcriptional regulators.

Because SRY and IRE-ABP share binding specificity for a sequence in a glycolytic gene, it is possible that these proteins may have similar or complementary functions. We have detected IRE-ABP messenger RNA in lipogenic tissues (fat and liver; unpublished observations) in which insulin regulates GAPDH gene expression, and in cultured Sertoli cells where metabolism of glucose to lactate is very rapid^{14,15} in order to support spermatogenesis^{16,17}. Parallels may exist between the roles of SRY and a4/IRE-ABP in regulating metabolism and growth in adult tissues and in promoting differentiation of embryonic tissues, and also in the mechanism(s) by which they are regulated by extracellular signals. □

Received 11 July; accepted 3 October 1991.

- Nasrin, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5273-5277 (1990).
- Jantzen, H. M., Adman, A., Bell, S. P. & Tjian, R. *Nature* **344**, 830-836 (1990).
- Gubbay, J. et al. *Nature* **346**, 245-250 (1990).
- Sinclair, A. H. et al. *Nature* **346**, 240-244 (1990).
- Berta, P. et al. *Nature* **348**, 448-450 (1990).
- Jager, R. J., Anvret, M., Hall, K. & Scherer, G. *Nature* **348**, 452-453 (1990).
- Singh, H., LeBowitz, J. H., Baldwin, A. S. & Sharp, P. A. *Nature* **52**, 415 (1988).
- Staudt, L. M. et al. *Science* **241**, 577-580 (1988).
- Koopman, P., Munsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. *Nature* **348**, 450-452 (1990).
- van de Wetering M., Oosterwegel, M., Dooijes, D. & Clevers, H. *EMBO J.* **10**, 123-132 (1991).

- Waterman, M. L., Fischer, W. H. & Jones, K. A. *Genes Dev.* **5**, 656-669 (1991).
- Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. *Genes Dev.* **5**, 880-894 (1991).
- Waterman, M. L. & Jones, K. A. *New Biologist* **2**, 621-636 (1990).
- Robinson, R. & Fritz, I. B. *Biol. Rep.* **24**, 1032-1041 (1981).
- Jutte, N. H. P. M. et al. *J. Rep. Fert.* **65**, 431-438 (1982).
- Labrie, F. & Proulx, L. (eds) *Endocrinology* 615-630 (Elsevier, Amsterdam, 1984).
- Skinner, M. K. *Endocr. Rev.* **12**, 45-77 (1991).

ACKNOWLEDGEMENTS. We thank I. Herskowitz, D. D. Moore and J. Avruch for discussion and critical reading of the manuscript; L. Ercolani for access to nucleic acid data bases; P. Banerjee for a rat adipocyte library; L. Gierle, M. Denaro and G. Respress for technical assistance, and C. Hughes for secretarial assistance.

Novel guanosine requirement for catalysis by the hairpin ribozyme

Bharat M. Chowrira, Alfredo Berzal-Herranz & John M. Burke*

Department of Microbiology and Molecular Genetics, Markey Center for Molecular Genetics, The University of Vermont, Burlington, Vermont 05405, USA

THERE is much interest in the development of 'designer ribozymes' to target destruction of RNAs *in vitro* and *in vivo*¹. Engineering of ribozymes with novel specificities requires detailed knowledge of the ribozyme-substrate interaction, and a rigorous evaluation of sequence specificity. The hairpin ribozyme catalyses an efficient and reversible site-specific cleavage reaction²⁻⁴. We have used mutagenesis and *in vitro* selection strategies to show that RNA cleavage and ligation has an absolute requirement for guanosine immediately 3' to the cleavage-ligation site. This G is not required for efficient substrate binding, rather, its 2-amino group is an essential component of the active site required for catalysis.

We tested the binding and cleavage activity of the hairpin ribozyme (Fig. 1a) against a panel of 12 variant substrates (Fig. 1b), representing all possible single-base changes in the four unpaired bases surrounding the cleavage site. The ribozyme cleaved all nine substrate molecules containing single-base substitutions of A₋₁, U₊₂ and C₊₃ with reasonable efficiency. Six of these variants (A₋₁U, A₋₁G, A₋₁C, U₊₂A, U₊₂C and C₊₃U) were cleaved to essentially the same extent as wild type. Three (U₊₂G, C₊₃A and C₊₃G) were cleaved to a lesser extent than wild type (68%, 66% and 70%, respectively). Steady-state kinetic analyses (Table 1) of the most poorly cleaved variants confirmed that base substitutions at position -1 bring about only minimal reductions in catalytic efficiency (k_{cat}/K_M), whereas some but not all substitutions at +2 and +3 reduce efficiency by 10- to

20-fold. These data suggest that U₊₂ and C₊₃ may be involved in tertiary contacts that facilitate, but are not essential for, proper active-site alignment. But bases at position -1 are probably not recognized by the ribozyme.

By contrast, the ribozyme was unable to cleave any of the three substrate variants containing G₊₁ substitutions (G₊₁A, G₊₁U, G₊₁C, Fig. 2a). No cleavage products were detected after prolonged incubations (36 h), high ribozyme concentrations (fivefold molar excess of ribozyme over substrate), high Mg²⁺ concentrations (100 mM), varying formamide concentrations (0.03 M-3.0 M), and temperatures (4-50 °C). On the basis of results with poorly cleaved substrates⁵, we estimate that the efficiency of cleavage of the G₊₁ variants is decreased by a factor of at least 10⁵.

We applied an *in vitro* selection method, recently developed in our laboratory (A.B.-H., S. Joseph and J.M.B., unpublished results), independently to examine specificity. This method permits the simultaneous analysis of many multiple sequence variants, and so complements the analysis of single-base substitutions that is the focus of the work presented here. After randomization of positions -1 to +3 to give a starting pool of 256 variants, molecules active in both cleavage and ligation were identified. Of the 33 active molecules examined, all molecules contained G₊₁, whereas none of the 15 inactive molecules examined contained G₊₁ (data not shown). No significant bias was seen at other positions in either the active or inactive pools. These results demonstrate that all molecules lacking G₊₁ are inactive, and strongly suggest that all variants in the substrate portion of the internal loop surrounding the cleavage site that contain G₊₁ can be cleaved and ligated by the ribozyme.

Why is the ribozyme unable to cleave G₊₁ variants? Two fundamentally different models would explain these results. Base substitutions might prevent substrate binding. Alternatively, they could interfere with an essential step after binding. Binding assays and inhibition experiments showed that variants G₊₁A, G₊₁C and G₊₁U were efficiently bound to the ribozyme (Fig. 2, Table 1). Dissociation constants (K_d) for binding of the

* To whom correspondence should be addressed.

TABLE 1 Kinetic and binding parameters for substrate variants

Substrate*	Sequence*	K_M (nM)†	k_{cat} (min ⁻¹)†	Relative k_{cat}/K_M	K_i (nM)†	K_d (nM)‡
Wild type	UGACA↓GUCCUGUUU	30 (±4)	2.0	1.0	—	—
A ₋₁ C	UGAC↓GUCCUGUUU	30 (±8)	1.1	0.6	—	—
G ₊₁ A	UGACA↓AUCCUGUUU	—	0	0	77 (±5)	47
U ₊₂ G	UGACA↓GGCCUGUUU	110 (±16)	0.5	0.06	—	—
C ₊₃ A	UGACA↓GUACUGUUU	58 (±29)	0.4	0.1	—	—
Variant	UÇACA↓GUCCUGUUU	32 (±15)	1.8	1.0	—	—
Variant (2AP)	UÇACA↓2APUCCUGUUU	65 (±14)	2.0	0.5	—	—
Variant G ₊₁ A	UÇACA↓AUCCUGUUU	—	0	0	—	33
Variant G ₊₁ I	UÇACA↓IUCCUGUUU	—	0	0	—	61

* Underlined nucleotides differ from the wild-type substrate. Variant, substrate for variant ribozyme that contains a single G at the cleavage site⁵ (see legend to Fig. 3). Variant and variant (2AP) were synthesized by solid phase RNA phosphoramidite chemistry^{5,18}. All other oligonucleotides were generated by transcription of DNA oligonucleotides using T7 RNA polymerase, and contain the sequence GCG (ICI) at their 5' ends in addition to the bases shown.

† Michaelis-Menten analysis was done as described⁵ in standard cleavage buffer at 37 °C. Analysis of inhibition using the noncleavable substrate analogue G₊₁A was carried out using inhibitor concentrations of 0.3 and 1 μM, and showed that G₊₁A is a competitive inhibitor.

‡ Dissociation constants were calculated from data such as those in Fig. 2b and as described previously¹⁶.

G₊₁A substrate analogues by wild-type and variant ribozymes correlate extremely well with the value for K_i obtained from kinetic analysis of ribozyme inhibition and correspond to binding energies (ΔG°) for substrate analogue binding of -9 to -10 kcal mol⁻¹. Values of K_d correspond well to previously reported values for K_M ⁵.

Collectively, these results lead to the conclusion that G immediately to the 3' side of the cleavage site is necessary for cleavage and ligation, although the base itself does not contribute significantly to substrate binding by the ribozyme. Thus, the G requirement is manifested at a step after substrate binding.

What is unique about G at position +1? Molecular modelling studies indicated that, among the four naturally occurring bases, G alone could bring a chemically useful functional group into close three-dimensional proximity to the bond that is cleaved (Fig. 3). Specifically, if G₊₁ is in a *syn* conformation, its 2-amino group may lie close to the scissile P-O bond and to the 2'-OH of A₋₁, also known to be essential for cleavage^{5,6}.

To test the possibility that the 2-amino group of G might be involved in active-site chemistry, we examined the ability of the ribozyme to cleave substrates containing the G analogues inosine and 2-aminopurine, using a variant ribozyme that cleaves a

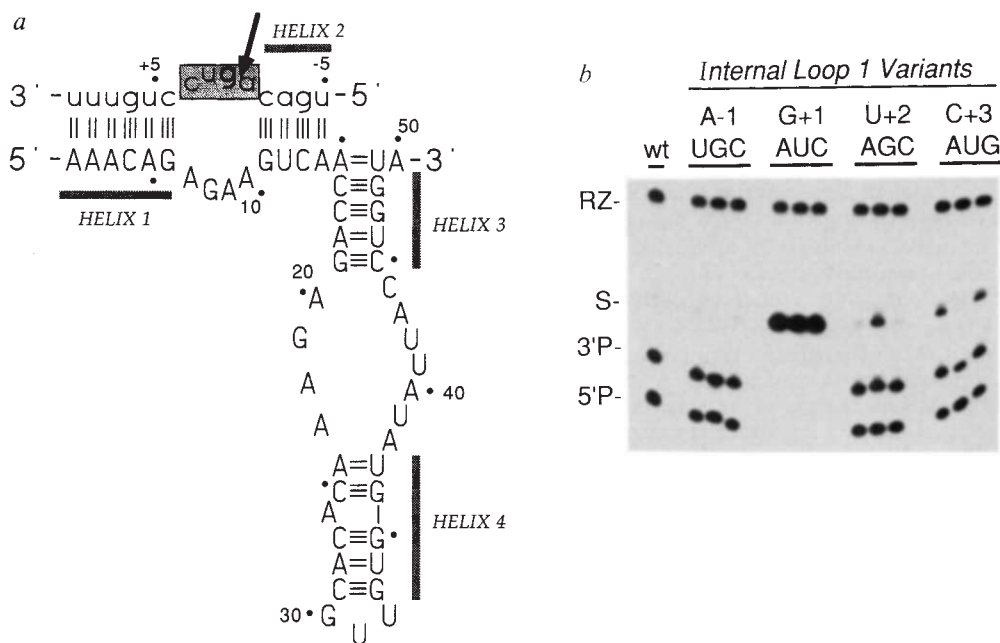
substrate containing a single G (ref. 5). Results showed that a substrate containing 2-aminopurine at the cleavage site was efficiently cleaved, whereas an inosine-containing substrate could not be cleaved (Fig. 3). The inosine analogue was bound with a K_d of 61 nM, a value comparable to that observed for binding of the G₊₁A substrate analogue by the variant ribozyme (Table 1). These results indicate unambiguously that the 2-amino group of G₊₁ is required for substrate cleavage and possibly ligation.

The 2-amino group of G₊₁ is likely to be an essential component of the active-site structure of the hairpin ribozyme. By contrast, the keto group of G₊₁ (absent in 2-aminopurine), is not implicated in catalysis. An alternative interpretation, that the 2-amino group of G₊₁ forms an important contact with the ribozyme that is not itself part of the active site, cannot be ruled out. But it is extremely unlikely that destabilization of such a contact could result in a $\geq 10^5$ -fold reduction in activity.

What is the mechanistic basis for the guanosine requirement of the hairpin ribozyme? It is possible that the 2-amino group serves as a general base, to facilitate deprotonation of the attacking 2'-hydroxyl at position -1 during cleavage. Alternatively, hydrogen bonding between the 2-amino group and a

FIG. 1 a, Hairpin ribozyme secondary-structure model. Upper case, ribozyme sequences. Lower case, substrate sequences. Arrow indicates substrate cleavage site. Substrate nucleotides to the 3' side of the cleavage site have positive numbers; those to the 5' side have negative numbers. Evidence for helices 1-4 derives from computer-assisted RNA folding, phylogenetic analysis, analysis of compensatory mutations and modelling^{2-4,13,14}. The boxed region indicates the region of the substrate analysed (internal loop 1). A variant ribozyme that cleaves a substrate containing a single G (ref. 5) was used and contains two substitutions (C₄ and C₁₃ changed to G). Its cognate substrate also contains two changes (G₊₆ and G₋₄ changed to C). b, Cleavage of variant substrates. RZ, ribozyme; S, 17-base substrate RNA; 3'P and 5'P, 3' and 5' cleavage products; Wt, wild-type substrate. Sequences of variants at positions A₋₁, G₊₁, U₊₂, and C₊₃ are indicated. α -³²P-labelled RNAs were synthesized using synthetic DNA templates as described^{5,15}.

Cleavage reactions were carried out for 30 min at 37 °C in a buffer containing 2 mM spermidine, 40 mM Tris-HCl (pH 7.5), and 12 mM MgCl₂ as described⁵. Reactions contained 40 fmol ribozyme and 200 fmol substrate in a volume



of 10 μl. Autoradiogram of 20% denaturing polyacrylamide gel is shown. The dried gel was quantitated by radioanalytic scanning using a Betagen instrument.

FIG. 2 Complex formation with noncleavable substrate analogues. *a*, Binding of $G_{+1}U$ and $G_{+1}C$ variants to the wild-type ribozyme. G_{+1} , unbound substrate G_{+1} variant; $G_{+1}:RZ$, ribozyme-substrate complex. The binding studies were done as described previously^{5,16,17}. Briefly, 20 nM ^{32}P -labelled substrate was incubated alone (-) or with 30 nM unlabelled ribozyme (+) on ice for 2 h in standard cleavage buffer. The complexes were resolved on a 10% native polyacrylamide gel at 4 °C containing 12 mM magnesium acetate as described⁵. *b*, Equilibrium binding of substrate variant $G_{+1}A$ to the ribozyme in 12 mM $MgCl_2$. Concentrations of unlabelled ribozyme (RZ) are indicated. Concentration of variant substrate analogue $G_{+1}A$ was held constant at 13 nM. $G_{+1}A:RZ$ is the labelled substrate-unlabelled ribozyme complex.

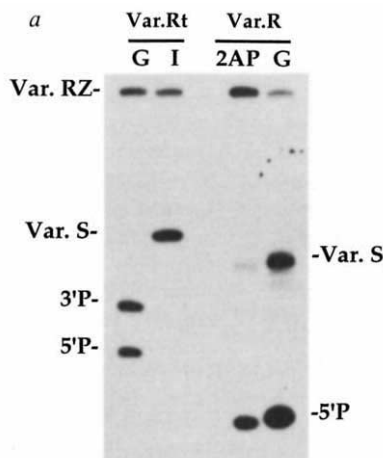
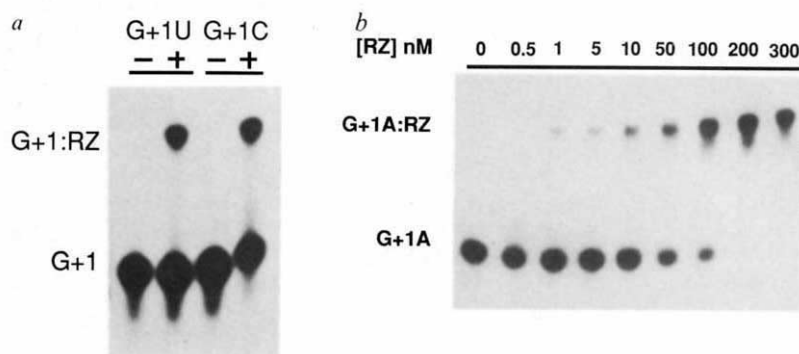
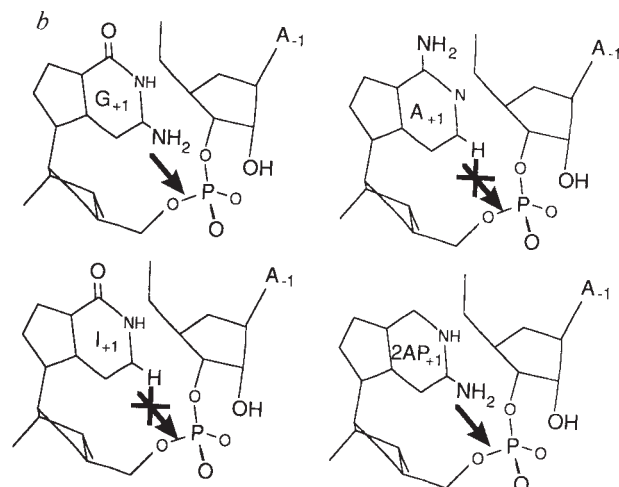


FIG. 3 *a*, Cleavage assays for inosine- and 2-aminopurine-containing substrates. For these experiments, a variant ribozyme (Var.RZ) that cleaves a substrate containing only a single guanosine (G_{+1}) was used. This ribozyme cleaves its cognate substrate (Var. S) with kinetics comparable to those of the wild-type ribozyme against its cognate substrate⁵. *G*, Variant G_{+1} guanosine substrate; *I*, variant G_{+1} inosine substrate; 2AP, variant G_{+1} 2-aminopurine substrate; Var. Rt, variant substrate RNA synthesized by transcription; Var. R, variant 14 nucleotide substrate synthesized by RNA phosphoramidite chemistry as described⁵; 3'P and 5'P, 3' and 5' cleavage product, respectively. Only one cleavage product (5'P) is seen under Var. R substrates because they are 5' end-labelled. Substrate containing inosine at +1 was synthesized by transcription in the presence of UTP, ATP, CTP and ITP (U.S.B.). T7 RNA polymerase can efficiently use ITP for chain initiation

phosphoryl oxygen could stabilize the transition state. Coordination of an essential magnesium ion at the active site by this moiety is unlikely as amino groups are poor ligands. It is interesting that the two functional groups essential for substrate cleavage and ligation by the hairpin ribozyme (2'-hydroxyl of A_{-1} and 2-amino group of G_{+1}) are both within the substrate, rather than the ribozyme.

The guanosine requirement of the hairpin ribozyme is unique. The hammerhead ribozyme does not require G at the cleavage site^{7,8}. Both the self-cleaving genomic and antigenomic strands of hepatitis delta virus have G to the 3' side of the cleavage site. But efficient cleavage of both genomic and antigenomic hepatitis delta sequences can occur when this G is mutated (M. Been



and elongation. Substrate containing 2-aminopurine at position +1 was synthesized chemically using 2-aminopurine phosphoramidite¹⁸. Cleavage reactions were done as described above, using a substrate:ribozyme ratio of 10. Ribonuclease T1 was used to confirm the presence of inosine and 2-aminopurine in the substrate RNA (data not shown). The Var. R substrates were tested for chemical purity by complete alkaline and RNase T2 hydrolysis (data not shown) as described⁵. *b*, Guanosine-mediated substrate cleavage. A model for the possible structure of the substrate cleavage site for the four bases and base analogues tested is shown. The arrows point to the scissile bond, and are not intended to convey information concerning catalytic mechanism. X through arrow indicates no reaction. Molecular modelling was done using HGS physical models (Maruzen) and computer graphics (Nemesis software for Macintosh, Oxford Molecular).

and G. Dinter-Gottlieb, personal communications). Ribonuclease P cuts pre-transfer RNAs immediately to the 5' side of the mature 5' end. In most cases the base at this site, which forms the first base-pair of the acceptor stem, is G. But ribonuclease P can correctly process RNA molecules lacking G at this site⁹⁻¹¹. Group I introns have well documented requirements for guanosine, but reactions always occur to the 3' side of G (ref. 12).

Our results are important for ribozyme targeting. We conclude that targets should be chosen that include the sequence 5'-N₁GHY-3', where H=U, C, or A, and Y=C or U. For optimal cleavage efficiency, the target should include either U_{+2} or C_{+3} . □

Received 17 July; accepted 13 September 1991.

1. Cech, T. R. *J. Am. med. Assoc.* **260**, 3030-3034 (1988).
2. Feldstein, P. A., Buzayan, J. M. & Bruening, G. *Gene* **82**, 53-61 (1989).
3. Hampel, A. & Tritz, R. *Biochemistry* **28**, 4929-4933 (1989).
4. Haseloff, J. & Gerlach, W. L. *Gene* **82**, 43-52 (1989).
5. Chowrira, B. M. & Burke, J. M. *Biochemistry* **30**, 8518-8522 (1991).
6. van Tol, H. et al. *Nucleic Acids Res.* **18**, 1971-1975 (1990).
7. Koizumi, M. & Ohtsuka, E. *Biochemistry* **30**, 5145-5150 (1991).
8. Ruffner, D. E., Stormo, G. D. & Uhlenbeck, O. C. *Biochemistry* **29**, 10695-10702 (1990).
9. Green, C. J. & Vold, B. S. *J. biol. Chem.* **263**, 252-257 (1988).
10. Burkard, U. et al. *J. biol. Chem.* **263**, 2447-2451 (1988).
11. Carter, B. J. et al. *J. biol. Chem.* **265**, 7100-7103 (1990).

12. Cech, T. R. *A. Rev. Biochem.* **59**, 543-556 (1990).
13. Rubino, L. et al. *J. gen. Virol.* **71**, 1897-1903 (1990).
14. Hampel, A. et al. *Nucleic Acids Res.* **18**, 299-304 (1990).
15. Milligan, J. F. & Uhlenbeck, O. C. *Meth. Enzym.* **180**, 51-62 (1989).
16. Pyle, A. M., McSwiggen, J. A. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8187-8191 (1990).
17. Fedor, M. J. & Uhlenbeck, O. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1668-1672 (1990).
18. Doudna, J. A. et al. *J. org. Chem.* **55**, 5547-5549 (1990).

ACKNOWLEDGEMENTS. We thank J. Doudna and N. Usman for synthesis of the substrate containing 2-amino purine, G. Dinter-Gottlieb and M. Been for communication of unpublished results and P. Grabowski and L. Kirsebom for valuable discussions. This work was supported by the NIH. A.B.-H. was supported by a NATO Postdoctoral fellowship.

Songs in the key of life

J. S. Jones

Exons, Introns, and Talking Genes: The Science Behind the Human Genome Project. By Christopher Wills. *Basic Books*: 1991. Pp. 288. \$23 (United States), \$30 (Canada), £23.

The Lottery of Life: The New Genetics of the Future of Mankind. By Philippe Frossard. *Bantam*: 1991. Pp. 252. £14.99.

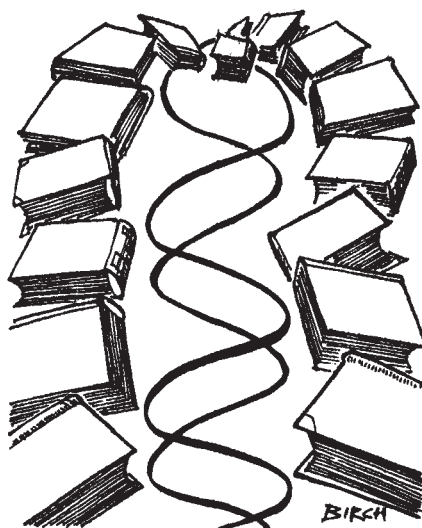
The Human Blueprint: The Race to Unlock the Secrets of Our Genetic Script. By Robert Shapiro. *St Martin's*: 1991. Pp. 432. \$24.95.

THE world's most boring book will be the complete sequence of the human genome: three-thousand-million letters long, with no discernible plot, thousands of repeats of the same sentence, page after page of meaningless rambling, and the occasional nugget of sense — usually signifying nothing in particular. The manuscript should be finished by the year 2005, and the first copy will cost a great deal — about the same, in fact, as a Trident nuclear submarine. All of us who have to explain to students just how long the genetic message is are forced to fall back on analogies: the genome is like so many copies of the *Encyclopaedia Britannica*, the *Oxford English Dictionary* or the Manhattan telephone book. Soon there will be another standard of comparison: putting all the popular works about the human genome project next to each other should, before long, form a shelf of volumes about the same length as the genetic message itself.

These three books are among the first of what will no doubt be an enormous *oeuvre*. Each is good in its own way, and each is impressively up to date (which means that each will soon be comprehensively out of date). They differ in emphasis, as do the expectations of those involved in sequencing. Wills is best: he gives a full, clear and attractively written account of the medical and ethical implications of the human genome project. Frossard (who works for a biotechnology company) is much concerned with its commercial implications, whereas Shapiro — a proponent of the 'little did he realize' school of historical writing — ranges widely through the past, present and future of genetics. Each has plenty of local colour, with interviews with many of those eminent in the field in their native habitats, although English readers may be baffled as to which moors Shapiro saw from Fred Sanger's laboratory in Cambridge.

There is a certain relentless optimism in all three. Each author assumes that it may one day be possible to cure genetic disease, ignoring much of the history of modern medicine, which has been far more successful at alleviating symptoms than producing cures. Wills, in particular, holds out high hopes for genetics: his

introductory chapter is a vivid account of how technology may soon help the brain-damaged child of a crack addict by persuading neurons to de-differentiate and regenerate themselves with the help of nerve growth factors — paying special



attention, of course, to those that increase IQ. Given that the United States, where this exciting future is promised, has a pattern of childhood mortality more characteristic of developing countries, one may wonder whether it would be more efficient to spend the money on improving a society that produces addicted children in the first place.

There is some interesting politics of science here too. The arguments among the government agencies involved in funding are spelt out in detail, together with some of the notably sordid commercial involvement. Frossard describes how Genentech offered stock to the chairman of the committee supervising a clinical trial, while his own company pursued the gene responsible for Alzheimer's disease largely because "the management was very greedy at the time". Although some fortunes were made — Robert Swanson and Herbert Boyer, who set up Genentech with \$1,000 capital in 1978, were each worth \$66 million by 1980 — biotechnology produced more South Sea Bubbles than Southern blots, with more than four-hundred companies founded in the United States, only a few of which

ever did anything useful.

The new genetics will produce plenty of ethical and legal problems. There are already quarrels about patenting segments of the human genome, about who owns the frozen embryos when a marriage breaks up, and about whether physicians are liable to pay damages to the parents for failing to predict the birth of an abnormal child. Whether the problems quite justify some of the more outrageous statements made by molecular geneticists is not clear. Irwin Chargaff, one of the founders of the subject, described the patenting of human genes as "a molecular Auschwitz, in which genes are extracted rather than gold teeth", while Watson himself (the one who appeared Before Crick) is quoted by Shapiro as threatening to aim nuclear missiles at Tokyo because the Japanese will not pay their fair share for sequencing the human genome.

The human genome project is a classic case of goal-directed research of the kind that put men on the Moon. Thirty years later, the general perception of the Apollo project is — so what? Genome research seems safe from such retrospective damnation. After all, it has already found the gene that makes me male, and is well on the way to finding the one that makes us all grow old and even, if you believe the optimists, to alleviating the symptoms of decay with a judicious dose of genetically engineered growth hormone.

The plan to sequence the human genome has been relentlessly hyped. There are some absurd predictions of what it can do. Those of Frossard, Shapiro and Wills are no exception. Once we have sequenced ourselves, we will know, apparently, which wine to buy, whether to learn a musical instrument and even whom to sleep with. The gene sequencers are pursuing the ultimate reductionist programme: to understand the message, we just have to put all the letters in order. There is an opposing view which suggests that, having sequenced the genome, we may be in the position of a nonmusician faced with the score of Wagner's *Ring* cycle: many pieces of information, apparently making no sense at all, but in fact containing an amazing tale — if only we knew what it meant. None of these books tells us this, but each of them is worth reading for what it says about the first bars of the first act of HUGO, the human genetic opera. □

J. S. Jones is in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK, and is delivering this year's Reith Lectures for the BBC.

■ *The Human Genome Project: Cracking the Genetic Code of Life* by Thomas F. Lee has just been published by Plenum, \$24.50.

Apes cast out of Eden?

W. C. McGrew

The Egalitarians — Human and Chimpanzee: An Anthropological View of Social Organization. By Margaret Power. Cambridge University Press: 1991. Pp. 290. £29.95, \$44.50.

OCCASIONALLY there comes a book that claims to change the whole picture in an area of science — that is, not just revise it but instead remake it. And when that book comes from one of the world's most important publishers of scholarship, it must be taken seriously.

Power says that most of what we know about the natural behaviour of our nearest living relation, the chimpanzee, is wrong. According to her, the two great field studies in Western Tanzania over the past 25 years are misleading and distorted. What has been reported by Jane Goodall and her colleagues at Gombe and by Toshisada Nishida and his colleagues in the Mahale Mountains must be junked. If Power is right, primatology is in a mess.

Her thesis is that chimpanzees studied in nature can be divided into two types: wild and provisioned. Wild apes are those "whose natural patterns of feeding and ranging have not . . . been changed by artificial feeding". Examples emphasized are early studies by Adrian Kortlandt and Vernon and Frances Reynoldses, and the early research at Gombe and Mahale before provisioning began. (Provisioning is giving supplementary, artificial food such as bananas or sugar cane to accelerate habituation; habituation is the progressive tolerance by wild animals to being watched at close-range by humans.) Thus, provisioned apes "are not confined but fed almost all of their food by people". This, Power says, applies to all the ethological research at Gombe and Mahale since the mid-1960s.

Thus the effect of provisioning is devastating; "artificial feeding is a catalyst for a high degree of qualitative, negative social change among the provisioned groups, which spread in ripple fashion to distort all aspects of the adapted social order". Limited food led to violence. The corrupted chimpanzees became xenophobic, murderous, cannibalistic, incestuous, infanticidal, despotic, retarded, anxious, greedy and so on. In short, sociopathy became the norm.

The demise of wild chimpanzees is all the more regrettable because, according to Power, they are really egalitarian. They live in friendly, open societies that are nonaggressive, nonterritorial and nonhierarchical. The sexes relate

through mutual dependence, and growing up is a matter of apprenticeship and acceptance. Child-rearing is collective, leadership is charismatic, sex is free, elders are wise, and the overall positive ambience is best exemplified in periodic 'carnivals', which are mass gatherings for socializing and reaffirmation.

The contrast could hardly be more stark, but it does not stand up. Field studies of chimpanzees cannot be so simplistically dichotomized; and neither Gombe nor Mahale actually fits Power's own definition of a provisioned population, because the artificial food provided only a fraction of the average daily intake of food. The heavy provisioning focused on by Power stopped long ago and was replaced by the giving of selective and intermittent snacks to individuals (not groups). Even this ceased years back, except in rare cases such as when medical intervention was called for. Power ignores this.



BETWEEN 2700 and 2400 BC, craftsmen in the Cycladic Islands of Greece sculptured these marble nudes whose elegance and simplicity were not seen again until the work of Brancusi and Modigliani. The sculptures range in size from under 10 cm to life size; their most striking feature is their adherence to a canon — most are female, standing, have folded arms, and lack surface detail. A comprehensive and stunningly illustrated account of Cycladic art is given by Colin Renfrew in *Cycladic Spirit*, published by Thames and Hudson, £32.

What of research on unprovisioned, wild chimpanzees? The early studies cited by Power were admirable pioneering efforts done when expectations and standards were different, and all had their limits. They were short (usually lasting months instead of years) glimpses of usually unidentified subjects of unknown age and status. The apes could be observed only opportunistically, usually at a distance. However hard-won, the sparse data yielded only suggestive hypotheses rather than firm conclusions. Only at Gombe and Mahale did the researchers push on to more sustained and intimate studies of behaviour in the daily lives of well-known individuals.

But the truest test of Power's reinterpretation of chimpanzee life in nature is provided by the wild populations of apes that show competition, hierarchy, territoriality, sex differences and so on, but that, crucially, have never been provisioned. Power either fails to mention these cases or consigns them to the fine print of her notes section. The most striking omission is that of the chimpanzees of the Tai forest of Ivory Coast studied by Christophe and Hedwige Boesch. In a series of articles published in mainstream journals during the 1980s, these authors described a chimpanzee society that, for example, competes forcefully for meat captured in frequent hunting forays.

So, why did Power get it wrong? She has sought to force chimpanzees into a Rousseau-like model borrowed from human foraging societies, depending heavily on Colin Turnbull and James Woodburn, while failing to acknowledge key differences. Unlike human gatherer-hunters, chimpanzees do not exchange their gains, and so are not mutually dependent in subsistence. Also, Power had tried to understand a complex species without ever having studied it or (apparently) without ever consulting anyone currently studying chimpanzees in the field. Finally, she has attempted to make sense of an intelligent, strategizing creature while largely ignoring the advances of modern evolutionary biology. Interpretations couched in terms of group or species selection have not proved fruitful for analysing vertebrate societies.

Provisioning is indeed a problematical field technique, as has been recognized for many years. It has rightly been largely abandoned. But it was unlikely to have been the Fall that cast out the apes from their equatorial Garden of Eden. Instead we must face up to the fact that chimpanzees are probably very like us, warts and all. □

W. C. McGrew is in the Department of Psychology, University of Stirling, Stirling FK9 4LA, UK.

Coming to terms with animals

Ralph A. Lewin

The Concise Dictionary of Zoology. By Michael Allaby. Oxford University Press: 1991. Pp. 508. £20, \$39.95.

If you are working at the cutting edge of zoology — which I conceive to be more like a fractal fringe — you may not find your pet animal listed here. After all, you may be studying it precisely because so little is known about it. For instance, my own work years ago concerned the *bursa Fabricii** of woodpigeons*: later it took me to coral-reef areas, some depredated by crown-of-thorns* starfish, others patchily carpeted with didemnids* (phycozoans* containing algae long regarded as zoochlorellae*) I found there a tectibranch* mollusc with green symbiotic chloroplasts from its ingested algae. My wife works with sea-skaters* (*Halobates**), perhaps the most widespread of all insects, and other components of the marine pleuston*. (Neuston is not the same as pleuston: neustonic organisms live under the surface, pleustonic ones on top of it.) None of the asterisked words appears in this dictionary.

It is a general dictionary practice to alphabetize nouns and not their qualifiers when listing such items as Andean condor, bee humming-bird, edible frog, imperial pigeon, lungless salamander, noisy scrub-bird, rock rat, rough-toothed dolphin, screech owl and spider-hunting wasp. Allaby does otherwise, which I consider a point in the dictionary's disfavour. But blue butterflies, with a special family identity, could reasonably be listed in this way, and no rational zoologist would look up blue-bottles as bottles, blue.

My wife and I spent a couple of hours thinking of words that could have been included but were not, and easily came up with some 60. *Acantharia*, alewife, appendicularia, bloodworm, bobcat, brolga, budworm, cahaug, caprellid We then checked to see which of these could be found in the two comparable dictionaries of zoology in our library: that by A. W. Leftwich published by Constable in 1973 (7,600 words) and that by S. W. Pennak published by Ronald in 1964 (19,000 words). About one-quarter of our items were found in Leftwich, and about two-thirds in Pennak. (We did not make such comparisons the other way round.) Our friendly librarian, who also compared these three dictionaries, expressed an enthusiastic preference for Allaby, which of course is much more up to date than the others. (Another colleague to whom I showed this dictionary dismissed it because it does not include the vocabulary of molecular biology. I think this

criticism is unreasonable: for such terms one should look elsewhere. Likewise, one should not expect to find here more than a small fraction of the 100,000 terms listed in one of the bigger medical dictionaries.) It was also pointed out to me that popular names of fishes in particular are far from standard. A monkfish may be an angler-fish in one diocese, a shark in another. Definitions of eels — common or garden or whatever — are even more slippery. The Fish List, a Food and Drug Administration guide published in 1988 by the US Government Printing Office, Washington DC, has many other examples of this sort.

With about 300,000 (give or take an order of magnitude) genera of animals, living and fossil, ranging in size from dinoflagellates to dinosaurs, one could not expect more than a small fraction to be listed here. This is primarily an English-language dictionary, with about 6,000 entries, and as such it should not be overlaid with Graeco-Latin names. Only a few-hundred generic names are included. I would have preferred to find a few more with special biochemical, economic, genetic, political or sociological importance, such as *Acanthaster*, *Achatina*, *Artemia*, *Betta*, *Caenorhabditis*, *Lumbricus* and *Tribolium*. (For fun, I would also like to have seen included a few items such as griffin, mermaid, rhinograde, roc, unicorn and yeti; but maybe this is too serious a compilation

for such frivolities.) There are many names of families (usually, although not consistently, with numbers of genera), orders (usually with numbers of families), classes and higher taxa. Also listed are popular as well as classical (unpopular?) names, ecological concepts, physiological phenomena and so on. Certainly some special names for funny bits of scales and spines, ossicles and follicles, peculiar lobes of mandibles and recondite bits of genitalia unmentionable here have been omitted, perhaps rightly. But on the whole, this is a fine compendium of unquestionable use. I have already found myself referring to it as I confer with my zoological colleagues or watch nature programmes on television.

What would I recommend for the next edition? (There surely will be one.) I would suggest excluding terms not immediately relevant to a zoological dictionary, such as acid, biosynthesis, chisquared, information statistic, innate, kinase, over-specialization, quinone secondary structure, Schiff's reagent and wobble hypothesis. And I would reduce the length of many of the entries, as this is not an encyclopaedia or a textbook. Do we need more than 100 words for entries such as Asilidae, thelytoky and thylacine? I think not: saving some of the space and including more different entries might make the book even more valuable next time round.

Rightly, I think, derivations of terms have not been given. Cross-indexing and proof-reading seem to have been done impeccably, as one would expect for a work from this publishing stable. Make sure you have an Allaby handy: you will undoubtedly find it useful. □

Ralph A. Lewin is at the Scripps Institution of Oceanography, University of California, La Jolla, California 92093-0202, USA.



In oriental mythology the toad is represented as a creature of supernatural power, but also viewed with affection and described with wit. The Asian *Bufo* is usually found presiding over magic, being a master of escape who takes pills that transform humans into toads (or vice versa) and who knows the secret of eternal life. *Bufo* is lumped together with the loathsome creatures that represent the "evil, dark feminine forces" of *yin*, the ultimate manifestation of which is the Moon; so it is hardly surprising that the Chinese see not a man but a toad in the Moon. In a story that dates to the Han dynasty (200 BC), the Moon-toad, apparently not satisfied with pills, attempts to swallow the

Moon itself, causing an eclipse. These and other such fascinating details can be found in Robert DeGraff's *Book of the Toad*, a colourful and witty pot-pourri on the toad's place in science, history and mythology. Published by Park Street/Lutterworth, price is \$19.95, £12.95.

The heroic art of storytelling

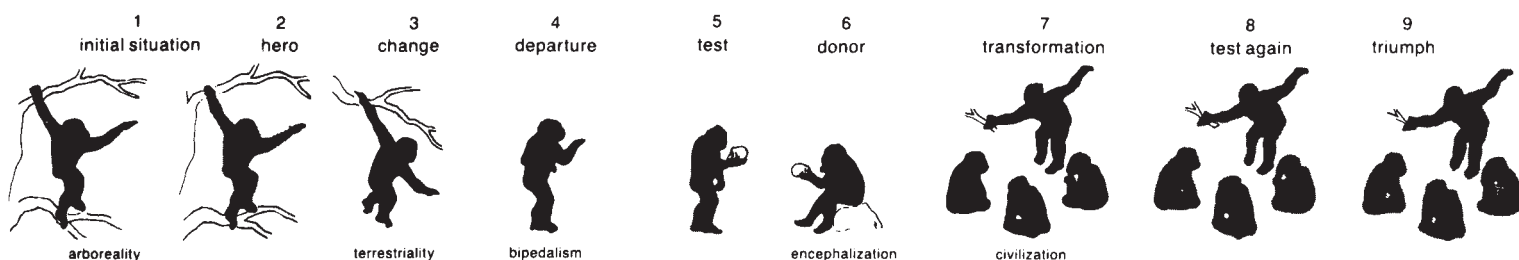
Glenn C. Conroy

Narratives of Human Evolution. By Misia Landau. Yale University Press: 1991. Pp. 202. \$25, £14.

SOME people get no respect. Thomas Macaulay once said that the more he read Socrates, the less he wondered why they poisoned him, and Thomas Jefferson was no less charitable to Plato when he remarked that the only thing remaining after one took away his "sophisms, futilities and incomprehensibilities" was

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

ultimately civilization. Obviously, for Darwin, principles of natural selection were the operative mechanism (the hero's helper or donor) during this journey. For Keith, on the other hand, orthograde posture evolved while ancestral humans were still in trees and this, rather than terrestriality *per se*, was the initial point of departure for human evolution. As an anatomist fascinated with the newly discovered role of hormones in growth and development, Keith subscribed to the orthogenetic view that "the machinery of evolution will be found inside the factory of the womb, rather than in the mechanism of natural selection". (Keith was an influential champion of the authenticity of the Piltdown Man and Galley Hill, and a



The sequence of events in Darwin's account of human evolution.

"his foggy mind". Fossil hunters, too, have been held in similar contempt by their detractors, often molecular biologists and cladists, who sometimes regard the literary byproduct of their sweat, blood and tears (to coin a phrase) as little more than the art of storytelling.

In *Narratives of Human Evolution*, the palaeoanthropologist-cum-literary-critic Misia Landau, rescues the art of anthropological storytelling by analysing just how complex these stories really are and what their underlying structures reveal about the basic evolutionary and social principles of their authors (what she refers to as "this altar housing a deep diversity of faiths").

Landau considers any account of a sequence of events that manifests "a deeper kind of belonging" (such as the story of human evolution) as a form of narrative that can be studied by techniques of literary criticism. Using methods originally developed by Vladimir Propp in *Morphology of the Folktale* (University of Texas Press, 1928), an analysis of Russian folktales, Landau first breaks down narratives of human evolution into their component, or functional, parts. A critical step in such an analysis is to dissect the tale "according to the functions of its *dramatis personae*" (that is, the evolutionary heroes). These functions then become the basic components of the tale.

Using a case-study approach, Landau liberally quotes from the works of the influential nineteenth- and twentieth-

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

century biologists Thomas H. Huxley, Charles Darwin, Ernst Haeckel, Arthur Keith and Grafton Elliot Smith in order to identify nine functions that she believes characterize all narratives of human evolution: (1) the initial condition in which the evolutionary hero is found (in all narratives of human evolution this setting is usually in a relatively carefree arboreal environment); (2) the hero is introduced as being somehow slightly 'different' from other nonhuman primates lurking in the vicinity; (3) the hero has to leave home (the change of situation in being dislodged from his arboreal home); (4) the hero departs to begin a new journey or adventure (for example, life as a terrestrial biped); (5) the hero is tested (by predators, climate, other

afarensis seems almost trivial in comparison to her more compelling and dramatic critical assessment of Darwin, Huxley and Haeckel's struggles with more herculean evolutionary themes. I would be fascinated to see Landau dissect contemporary icons of palaeoanthropology with the same surgical blades she uses on our idols of the past.

Science can be conveniently divided into three classes: the 'soft sciences' such as sociology and political science, the 'hard sciences' such as physics and chemistry and then the 'really difficult sciences' such as palaeoanthropology. *Narratives of Human Evolution* is a literate, thought-provoking explication of this really difficult science. I would add only one further thought to Landau's story. She notes that Huxley never did bridge that gap between the apes and humanity with a missing link. Maybe that is because the missing link is man. □

Glenn C. Conroy is in the Department of Anatomy and Anthropology, Washington University Medical School, St Louis, Missouri 63110, USA.

Theories for everything

A. Hallam

Extinction: Bad Genes or Bad Luck? By David M. Raup. Norton: 1991. Pp. 210. \$19.95.

The Miner's Canary: Unravelling the Mysteries of Extinction. By Niles Eldredge. Simon and Schuster: 1991. Pp. 246. \$20.

ALTHOUGH argument persists about the extent to which Darwin was a gradualist when it came to modes of speciation, with *The Origin of Species* being quoted like so much Holy Writ, there is no room for doubt about his views on the relative importance of biotic as opposed to physical causes in promoting extinction. "Species are produced and exterminated by slowly acting causes . . . and the most important of all causes of organic change is one which is almost independent of altered . . . physical conditions, namely the mutual relation of organism to organism — the improvement of one organism entailing the improvement or the extermination of others". This 'struggle for existence' view has been accepted uncritically by generations of evolutionary biologists, and is well expressed in Van Valen's famous Red Queen hypothesis, named after Lewis Carroll's character who found that, in her sort of country, it took all the running one could do to stay in the same place.

Thus, in the darwinian view, the biologically 'superior' mammals would have progressively outcompeted the dinosaurs through the course of the Mesozoic era, until dinosaurs eventually became extinct. It is now clear that this did not happen — mammalian radiations had to await the disappearance of the dinosaurs together with many contemporary terrestrial and marine organisms, a disappearance caused by mass extinction resulting from drastic environmental deterioration at the end of the Cretaceous period. Any competition between the dinosaurs and mammals must therefore have been pre-emptive rather than displacive, the key requirement being to be the first to occupy the ecological niche. This pattern of change appears to be characteristic of the fossil record as a whole, with episodic mass extinctions clearing the decks, as it were, for the radiation of new organic groups. In Raup's words, extinction appears to have been more a matter of bad luck than bad genes, because normal darwinian finely tuned adaptations would often have been a poor defence against a rare environmental catastrophe.

These two books on extinction by leading US palaeontologists are aimed at a wide audience and demand little technical knowledge. Both books have considerable merits but are very different in their treatment of the subject, reflecting the different aptitudes and interests of their authors. Raup's great strength is his skill in handling data statistically and his ability to make incisive points on the basis of this quantitative approach. He presents a penetrating discussion of randomness, based on such diverse topics as the 'gambler's ruin' problem and the skewed distributions ubiquitous in nature, and uses this to argue cogently for rarity being the main factor promoting extinction. He makes a pertinent comparison of mass extinctions with more familiar natural phenomena such as earthquakes and floods, with the events of greatest intensity being the least frequent. His disturbingly named "kill curve" depicts the average number of species killed for a series of waiting times, and hence tells us the average likelihood of a given event in a given length of time; needless to say, mass extinctions fall high on the curve.

The title of Eldredge's book refers to the canaries used by miners as early warning systems for poisonous gases. The recent decline of migrating songbirds fits the same role on a global scale, and is clearly a measure of habitat destruction, which in his view has always been the main cause of extinction. Eldredge writes as lucidly as Raup, but his style is less dry and more eloquent, and his book is richer than Raup's in biological examples. He outlines the

reasons for the diversity of the organic world, indicating that rates of speciation and extinction are higher in specialist forms: the higher diversity of the tropics is largely a consequence of finer niche-partitioning among more stenotopic organisms. Mass extinctions are in his view dramatic events that extended over millenia or longer, and are not normally the consequence of geologically instantaneous catastrophes.

In reviewing the possible physical causes of mass extinctions in the pre-human past, neither author is dogmatic and both freely admit to the continuing uncertainty and dispute within the scientific community. Because Raup believes that nature would have great difficulty eliminating species over large areas, he favours the crash of asteroids or comets into the Earth as the cause of not just mass extinction, but background extinction as well. This is, to my mind, a truly astonishing conclusion, being based on virtually no evidence apart from the pronounced iridium anomalies and shocked quartz at the Cretaceous/Tertiary boundary, which is the only reason why the impact story is taken seriously at all. Eldredge takes what I believe to be the geologically more reasonable view, for which there is much evidence from the stratigraphic record. In this view, important environmental changes bound up with events confined to our planet are responsible for mass extinctions. He sees no need to invoke *deus ex machina* of extraterrestrial events except for the Cretaceous/Tertiary boundary extinctions; even here, any impact was probably no more than a *coup de grâce* to elements of an already perturbed biosphere.

Although he does not discount sea-level change, Eldredge favours climatic deterioration as the most important extinction-inducing factor. I think that he overstates the case. Whereas falling temperature appears to have been implicated in the Cenozoic and late Ordovician extinctions, the evidence for other times ranges from poor or equivocal to nonexistent. Further, there is a strong correlation between marine extinction and global sea-level change: relatively sudden falls led to extensive regression of epicontinental seas, and, probably more importantly, sharp rises were associated with the spread of anoxic waters. Both could have caused devastating loss of habitat area. Despite the strong evidence for these phenomena, Raup does no more than mention them in passing, whereas Eldredge ignores them completely. In particular, these events get around the objection made by Raup that the pronounced regressions of the Quaternary period do not correlate with important extinction episodes, and hence the case

for sea-level change is not supported. The Quaternary world, with its extensive polar icecaps and well-oxygenated ocean, is an unreliable guide to most of Phanerozoic time.

It is clear that we are witnessing a mass extinction that is almost entirely the result of human activities. Eldredge is surely right that this is largely the consequence of habitat destruction rather than direct slaughter, but the consequences are nonetheless dire for

our own species as well as others. Conservation is more than just a matter of saving the odd squirrel species. To echo the words of a recent disenchanted US Secretary of the Interior, it involves the management of entire ecosystems, which falls in the sphere of politics as much as science. □

A. Hallam is in the School of Earth Sciences, University of Birmingham, Birmingham B15 2TT, UK.

Sceptical inquiries of old

Jeremy A. Sabloff

Fantastic Archaeology: The Wild Side of North American Prehistory. By Stephen Williams. University of Pennsylvania Press: 1991. Pp.432. \$28.95, £27.50 (hbk); \$14.95, £11.95 (pbk).

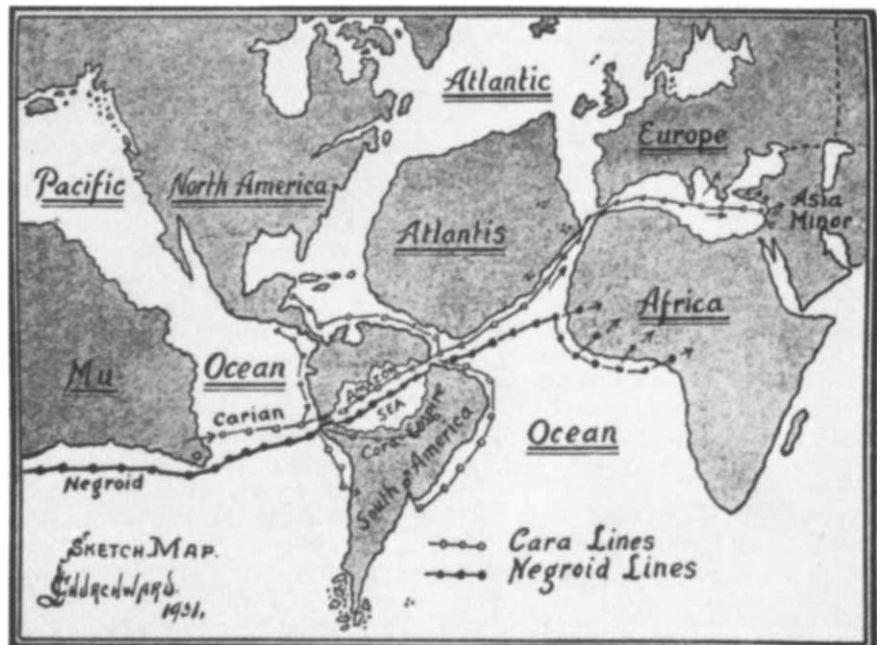
NEARLY three decades ago, the late Robert Wauchop, in his brilliant book *Lost Tribes and Sunken Continents* (University of Chicago Press, 1962), wrote that in response to the widespread popularity of the work of pseudoscientists, "the average professional anthropologist cannot or will not write the kind of book that people in great numbers will want to read". Fortunately, in the years since Wauchop made that damning statement, a growing number of practising scientists have entered the public fray. *Fantastic Archaeology* by Stephen Williams, Peabody Professor of North American Archaeology and Ethnology at the Peabody Museum, Harvard University, is a superb example of this recent trend. Engagingly written, the book's clear discussion of a variety of archaeological controversies is readily accessible to the general reader and should be of interest to both lay people and professional scholars alike.

More than any other discipline, archaeology has been unsuccessful in convincing the public at large that there is a clear line between the writings and conclusions of professional scholars and amateur dabblers. Although Williams does not closely examine the cultural reasons for this problem nor the huge

public appeal of archaeology in both its scientific and pseudoscientific forms, he does offer one of the best descriptions

origins of North American archaeology and the main intellectual debates of the nineteenth century, Williams forges on to examine a host of fascinating and controversial topics, including famous frauds such as the Grave Creek Stone, the Cardiff Giant and the Davenport Tablets; Constantine Rafinesque and the Walam Olum legend; claims for early Ice Age finds; the supposed lost islands of Atlantis and Mu and their advocates, particularly Ignatius Donnelly, Helena Blavatsky and the Theosophical Society, and James Churchward; archaeology and religion; alleged Viking finds such as the Kensington Stone and the Newport Tower; assertions of Asian influences on ancient North America with special attention to the writings of Harold Gladwin; the 'decipherments' of averred prehistoric inscriptions by writers such as Barry Fell; and the pronouncements of psychics such as Edgar Cayce, Jeffrey Goodman and Stephen Schwartz.

Williams concludes with a synopsis of



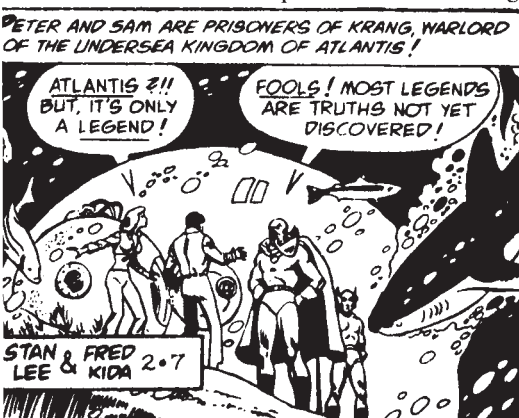
Sunken continents — Churchward's map of Atlantis and Mu, showing Mu's lines of influence on the whole picture of "man's advent on Earth", as he put it.

available of the differences between archaeological and fantastic accounts of various aspects of North American prehistory. He defines fantastic archaeology at the start as "those alternative views of the past that use data and interpretation that will not stand close scrutiny" and proceeds to offer a useful and tolerant overview of its nature and practitioners. The procedures of these purveyors of myth and fantasy are tellingly contrasted with those of trained archaeologists and the latter's use of the scientific method. In addition, the general criteria by which archaeological hypotheses (or assertions) of all stripes can be evaluated are concisely outlined.

Then, after a lucid discussion of the

what he calls "the real fantasy", the scientifically reconstructed history of the cultural achievements of the Native American inhabitants of North America. He maintains that the latter story has as many exciting and intriguing elements as the concocted histories of the pseudoscientific cranks. Archaeologists must strive to make the public aware of this story, Williams convincingly argues, while continuing to be "skeptical inquirers after knowledge and relentless foes of fraud and unreason". □

Jeremy A. Sabloff is in the Departments of Anthropology and the History and Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.



Getting to grips with the globe

Gordon L. Herries Davies

To Interpret the Earth: Ten Ways to be Wrong. By Stanley A. Schumm. Cambridge University Press: 1991. Pp. 134. £20, \$24.95.

"THE moment I hear the word philosophy I reach for my hammer." Such, I suspect, might be the sentiment of many a geologist, and for that reaction I am prepared to hazard an explanation. Geology likes to be perceived as a science abounding in machismo. This is clearly a delusion (worthy sisters have long since infiltrated 'the brethren of the hammer'), but many of us do still think of geologists as blokes who are burly and bearded, booted and bluff, beery and bawdy.

Hard-rock men

Certainly in the United Kingdom there is within the science itself a graded gender scale. Really masculine geologists grapple with the Palaeozoic era and older problems; those who dally with the Mesozoic and Tertiary periods are geologists only on sufferance; and those having affairs with the Quaternary are first branded as geomorphologists and then banished into the anile field of geography. Real British geologists are 'hard-rock men' (freudians will note the adjectives) and 'soft-rock men' feel just a teeny-weeny bit inadequate.

Now I imagine that most geologists, when pressed, would admit to regarding philosophy as an effete discipline — a discipline still more remote from the hard core of geology than is geomorphology — and it may be that many geologists imagine that their personal image would suffer through too close an involvement with philosophical issues. Here we may have an explanation for the fact that so many geologists go forth to do their geology with so little regard for the philosophical underpinnings of their science. Here, also, we perhaps have an explanation for the fact that many of those geologists who have reflected on the philosophy of their discipline have come from that extremity of the geological gender scale possessed of the least machismo — the subdiscipline of geomorphology. Names of such individuals that instantly spring to mind are G. K. Gilbert, W. M. Davis, T. C. Chamberlin, D. W. Johnson and J. H. Mackin. To those names there now has to be added that of S. A. Schumm.

Stanley Schumm has for several decades been one of the world's leading figures in the field of process geomor-

phology. With a wealth of practical field and laboratory experience behind him, he has now added to his *oeuvre* this broadly philosophical monograph intended to assist young Earth and environmental scientists standing on the threshold of a research career. It is a volume entirely without pretension. On the one hand, it disregards such important issues as the nature of geological truth or the rational basis for uniformitarianism, and Schumm freely admits that philosophers may find his volume somewhat lacking in depth. He certainly draws heavily on the work of such familiar figures as Beveridge, Chamberlin, Feyerabend, Gilbert, Kuhn, Medawar and Popper.

On the other hand, Schumm is not here concerned with methodological niceties at the level of the choice of maps suitable for exercises in morphometry or the principles that should guide the laboratory simulation of real-world processes. His objective lies midway between these two extremes and the heart of the volume consists of a parade of ten types of problem that may confront a student of the Earth's surface. Allusion to just three of these problems will serve to illustrate their genre.

There is the problem of 'location', which concerns the difficulty of extrapolating results from one observational site to another; there is the problem of 'convergence', when very different terrestrial processes yield similar effects; and there is the problem of 'singularity', where unexpected variation occurs between datasets. In discussing the ten problems, Schumm seeks to demonstrate that although there is no set of procedures that we may justly hail as embodying 'the scientific method', there is something that we may reasonably term 'the scientific approach'. It is this approach, Schumm suggests, that is the hallmark of the successful scientist; and it is this approach that Schumm here seeks to inculcate in his readers.

He presents his views both interestingly and effectively and the volume will serve as a useful vade-mecum for all those aspiring to an understanding of geomorphic processes. These individuals will turn to it for inspiration in moments of vacuousness and for guidance in moments of tribulation. The extensive bibliography may even encourage them to seek a deeper understanding not only of the character of their chosen specialism, but also of the very nature of science itself. Only once in these pages was my ire aroused. On page 56 the name of one of France's greatest geologists has become horribly *mutilé*. I refer to Alexandre Brongniart.

But if we study Schumm's words diligently, and apply his precepts rigorously, will we unfailingly arrive at

geological truth? As I read the volume, I carried in my mind the famed controversy that raged around 1800 over the origin of valleys. Which came first, the river or its valley? Were valleys ancient features and thus the progenitors of the occupying rivers or were they erosional features created by the resident rivers? Many eminent and experienced men favoured the former view, among them the Genevese geologist Jean André de Luc. It was de Luc who adduced a seemingly incontrovertible argument against the erosional hypothesis. If a valley contains a lake, he reasoned, then that lake is a natural settling basin for all the debris fluvially eroded from the valley upstream. Because the volume of sediment in such lakes is insignificant compared with the volume of the valley upstream, it would seem to follow that the valley upstream could never have been the work of the fluvial processes. De Luc died in 1817, more than 40 years before the advent of the glacial erosion theory inspired a recognition that the lake basins that had seemed so crucial are really of glacial origin and much younger than the containing valley. They have little if any relevance to the debate in which de Luc was involved. Now what I wonder is this: if Schumm's volume could have been placed in de Luc's hands in 1800, might its sound advice have rescued him and his *confrères* from their error? I think not. We are all of us prisoners of the age in which we live and our thought processes are shaped by the milieu to which we belong. We need to remember that knowledge of the Earth processes is not something that gradually seeps out of the Earth's surface to await the human harvester. Rather, that knowledge is created within fallible human minds as their owners gaze out upon the natural world.

Eleventh way

Schumm's volume will greatly assist all such observers in their reasoning, but at the same time we need to be mindful of our own inadequacies. That wise man John Ruskin in 1869 observed: "For the more I endeavour to read Nature patiently, the more I find that she is always trying to deceive us". Perhaps Schumm should have added to his monograph an eleventh source of error in our efforts to understand this Earth. That source of error is insurmountable. It is our humanity. □

Gordon L. Herries Davies is an emeritus fellow of Trinity College, University of Dublin, Dublin 2, Ireland.

■ In the second edition of *Continents in Motion*, Walter Sullivan presents a history of the idea of continental drift. AIP/Adam Hilger. Price \$50, £34.50 (hbk); \$25, £17.50 (pbk).

New in paperback

■ In *The Rebirth of Nature: The Greening of Science and God*, Rupert Sheldrake traces the mythological and historical roots of different attitudes to nature. James Lovelock described the book as "a new window into biology" (*Nature* **348**, 365; 1990). Published by Rider, £7.99.

■ Frank McKinney and Jeremy Jackson argue in *Bryozoan Evolution* that the growth pattern and form of a bryozoan colony is more an expression of its ecological niche than of its phylogenetic history. Published by the University of Chicago Press, price is \$15.95, £12.75.

■ *Weather Facts* by Dick File is a complete guide to the weather and climate based on the author's columns in the *Guardian*. Published by Oxford University Press, £6.99.

■ *Molecules* by P. W. Atkins is the most recent of the Scientific American Library series to be issued in paperback. Beautifully written and illustrated, the book is published by W. H. Freeman, price £10.95.

■ *Genome* by Jerry E. Bishop and Michael Waldholz is an accessible account of the science and motivations behind the human genome project. Published by Simon and Schuster, price £8.99. Reviewed in *Nature* **348**, 205 (1990).

■ *Lucy's Child: The Discovery of a Human Ancestor* by Donald Johanson and James Shreeve is a gripping review of latest thinking about human origins. Published by Penguin, price £6.99. Reviewed in *Nature* **348**, 373 (1990).

■ *The Hundreth Monkey* edited by Kendrick Frazier is the latest collection of essays from the *Skeptical Inquirer*. It includes contributions from Carl Sagan, Issac Asimov and Martin Gardner. Published by Prometheus Books, price £11.50.

■ *To Infinity and Beyond* by Eli Maor is a fascinating history of the infinite in mathematics, literature and art. Published by Princeton University Press, \$16.95, £15.

■ In *What Little I Remember*, Otto Frisch reminisces about the characters and events that shaped the development of modern physics. Published by Cambridge University Press, price is £5.95, \$10.95.

■ Three books originally published in AIP's *Masters of Modern Physics* series are now issued in paperback. They are: *The Charm of Modern Physics* by Sheldon L. Glashow (reviewed in *Nature* **351**, 706; 1991); *Citizen Scientist* by Frank von Hippel; and *The Road from Los Alamos* by Hans A. Bethe. Published by Simon and Schuster, price £8.99 each.

■ *Design for the Real World: Human Ecology and Social Change* by Victor Papanek has been translated into 23 languages since it first appeared in 1971; it has become the world's most widely read book on design, and is essential reading for those interested in aid, trade and technological development. The second edition is now published in paperback by Thames and Hudson. Price £10.95.

■ Volumes 1 and 2 of the highly acclaimed *A Course in Mathematics for Students of Physics* by Paul Bamberg and Shlomo Sternberg are newly published in paperback. Cambridge University Press, price £15.95, \$27.95.

Lawyers, lies and liability

James Randi

Galileo's Revenge: Junk Science in the Courtroom. By Peter W. Huber. *Basic Books*: 1991. Pp.274. \$23.

As an 'expert witness' on legal matters related to claims of alleged occult phenomena, I have often testified in court about specific and general findings in the field. So I thought I knew a bit about uninformed jurists and the rejection by courts of good evidence. Then I read *Galileo's Revenge*. What scales remained fell from my eyes, and I began to see the disturbing extent to which the scales of justice indeed tilt.

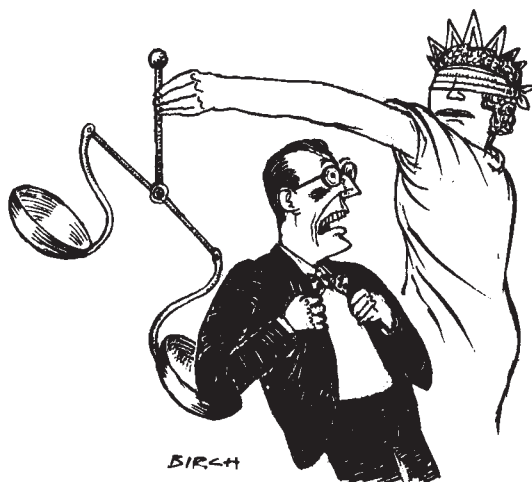
In this startling and riveting book, Huber reveals just how capricious the law is on the subject of liability, and how the outcome of a case can depend on the declarations of a claimed expert scientific witness who has been handsomely paid to have an opinion that serves his employer, a lawyer. At one time in the United States, such expert witnesses had to prove themselves by giving their credentials or by undergoing an examination. No longer. Law books actually instruct lawyers how to shop around. These so-called experts advertise, and agencies who supply them for employment state that if the expert they provide does not have the 'correct' opinion, then they will find the client one who does.

In 1986, a US federal judge advised his colleagues to "take hold of expert testimony in federal trials". As a result, some jurists began to pay some attention to what Huber calls "that radical London novelty of 1660 — the professional society — and that radical Paris novelty of the 1820s — the professional journal and peer review". But 'junk science', as Huber labels it, continued to rule, and still rules, in US courts. Untruths and raw speculation repeated enough times to jurists become the reality on which they base their decisions, leading to incredible judicial errors.

It is bad enough that the law is riddled with pseudoscientific notions; it is worse still when judges simply decide to ignore the best, established findings of good science. As an example, a California court ruled that Charles Chaplin was the father of a particular child despite the hard fact that Chaplin had group O blood, and the mother and child had group A and B blood, respectively. Unless a divine exception was made in this

case, Chaplin was not the father. But the court freely disregarded the scientific evidence.

Huber examines some of the most publicized legal controversies of recent decades. The frightening revelations that he provides range from the farcical "sudden acceleration" charges levied against the Audi automobile to the nonsense of "chemically-induced AIDS" claims. I must admit that I had accepted the claims made against the Audi model 5000 manufactured between 1978 and 1986. The company was forced to recall 250,000 cars that were perfectly sound,



and paid millions of dollars to drivers who simply mistook the accelerator pedal for the brake. Audi spent additional millions of dollars to defend themselves in court, and their sales in the United States dropped precipitously. My acceptance of the possibility that the car's computer habitually malfunctioned in a sporadic fashion was based on an account offered by a usually reliable television show that featured experts whose credentials were not easily checkable. I did not fall for the "environmental causes" claptrap that was invoked as a basis for litigation over AIDS. (Some cases were claimed to have been triggered from pollutants arising from carpets and furniture.)

What are the reasons for this deplorable state of affairs? In tackling this question, Huber looks at how and why even reputable scientists can slip into junk science, tracing as an early example the acceptance and subsequent rejection by the scientific community of the existence of N rays at the turn of this century. He shows how lawyers are now capable of launching massive attacks involving thousands of claims and extortionate settlements, as well as how the lure of insurance attracts and sustains litigation. Of course the mass media have a lot to answer for in creating enormous controversies with no

basis in fact. Indeed, as Huber notes, fear itself has become a basis for litigation.

Galileo's Revenge came out too soon to record two recent victories over junk science. Earlier this year, a woman in Florida claimed that she had lost her psychic powers when some timber fell on her at a local store. Although the judge declared before the case got under way that psychic powers had been scientifically established (a chilling pronouncement for the defence lawyers), the jury was more sensible and awarded her \$5,000. She had asked for \$1 million and had been offered \$40,000 to settle out of court. Sanctions against her took away the award and put her in debt to the court. Then in September, a federal jury absolved the Lilly Company from claims that their drug Prozac had caused users to commit murder and suicide, and had caused one woman to be an "insatiable nymphomaniac". The anti-Prozac cause had been supported by the Church of Scientology. In both these cases, junk science was rejected and reason prevailed, in spite of evidence given by paid 'scientific experts'.

Huber says that if an expert authority declares that something is in fact not the cause of a grievance, all the lawyers need to do is "ask a jury. And then another, and another, as many times in succession as the trial bar may deem to be justified by either visceral conviction or speculative greed. We find, once again, that our modern liability system is all accelerator and no brake." He is justifiably angry at jurists who deny the established scientific facts and accept pseudoscience: "It is simply unacceptable for any judge to insist that there is no such thing [as scientific truth] Claims dressed up in the form of serious science but lacking serious empirical and conceptual credentials will continue to be junk science. We do not hesitate to denounce charlatans and frauds when they peddle snake oil in country fairs or on network television. We need not hesitate to denounce them when they are primed and primed by lawyers and solemnly ushered into court." He contends that the rules of evidence must be changed to reflect science's definition of cause as the only one that is objectively verifiable.

We would do well to remember Huber's opinion when advocates of cold fusion and homeopathy object to the publication of valid commentary about the evidence for their claims. □

James Randi is a professional magician and investigator of pseudoscientific claims. He is the author of The Truth about Uri Geller, The Faith Healers and The Mask of Nostradamus, and holds a MacArthur Foundation Fellowship.

Intellectual dissent

Michael Shepherd

Final Analysis: The Making and Unmaking of a Psychoanalyst. By Jeffrey Masson. HarperCollins/Addison-Wesley: 1991. Pp. 212. £15, \$18.95.

PSYCHOANALYSIS has proved to be an important influence on Western thought during the twentieth century. Now that it is in decline, the reasons for its long run can be discerned more clearly. Scientific validity can barely be included among them. As Peter Medawar pointed out, the system is constructed on a quasi-scientific psychology; the ideas are essentially unbiological; the thinking is mythological, more akin to imaginative literature than to science; and the doctrines are "cunningly isolated from the salutary rigours of disbelief". Further, it has failed to justify the therapeutic claims advanced by its proponents.

Jeffrey Masson would agree, but from a very different standpoint. Most of the critics of psychoanalysis have examined it as informed outsiders. Masson has been an insider, a participant-observer who briefly entered the sanctum sanctorum of the psychoanalytical hierarchy when he was appointed projects director of the Sigmund Freud Archives by Kurt Eissler and formed a close professional relationship with Anna Freud. A former professor of Sanskrit, Masson rose rapidly within psychoanalytical circles during the 1970s and edited and translated the Freud-Fliess letters. Gradually,

however, he began to entertain doubts about several aspects of the world in which he appeared to be so successful, and in 1981 an article in the *New York Times* he gave public expression to his critical assessment of Freud's seduction theory, one of the cornerstones of psychoanalytical dogma. In the process he accused Freud of moral cowardice. Shortly afterwards he was relieved of his directorship of the archives and his links with the International Psychoanalytical Association were severed.

Masson has repented with his pen. In 1984 he published his views in a book with the self-explanatory title of *The Assault on Truth: Freud's Suppression of the Seduction Theory*. In 1988 he turned against the whole psychotherapeutic enterprise with another volume, again with a revealing title, *Against Therapy: Emotional Tyranny and the Myth of Psychological Healing*. This third volume is largely autobiographical, tracing the steps of the professional and personal journey that has led to his present position.

The story has been told before. In the United States, where psychoanalysis flourished as nowhere else, Masson's career was of sufficient interest for Janet Malcolm, a journalist on the *New Yorker*, to make it the springboard of her memorable exposé, entitled *In the Freud Archives* (Cape, 1984). Malcolm, whose earlier investigation of psychoanalysis as "an impossible profession" had familiarized herself with the background, provides a searching account of the issues and the people involved, including a vivid portrait of Masson himself, a man who "gave off the sheen of the intellectual big time that even those who disliked him from the start were impressed by".

The sheen is largely missing from *Final Analysis*, in which Masson presents himself more modestly as a man who looked for truth and was disillusioned by what he found, a man who was more sinned against than sinning and who gradually came to realize that he had landed in a nest of vipers. Making use of a succession of anecdotes and comments, he castigates the closed societies of the analytical institutes and depicts most of his former colleagues as venal, backbiting, envious and all too often incompetent. From this judgement he exempts Kurt Eissler and Anna Freud, who emerge as sincere but misguided fanatics. His verdict is unequivocal: "In my experience, psychoanalysis demanded loyalty that could not be questioned, the blind acceptance of unexamined 'wisdom'. It



Freud — accused of moral cowardice.

The Huber-Deutscher Collection

is characteristic of religious orders to seek obedience without scepticism, but it spells the death of intellectual inquiry. All variants of 'because I say so', or because the Koran says so, or the Bible says so, or the Upinashads say so, or Freud says so, or Marx says so, are simply different means of stifling intellectual dissent."

This is the voice of a disillusioned apostate, about whom several questions remain unanswered. The significance of his story, however, resides more in what it reveals of his former creed than in what it reveals of the author himself. He has provided a striking and well-publicized illustration of Ernest Gellner's dissection of the psychoanalytical movement as a highly organized secular faith, a complex belief system applying a scientific idiom to a variety of social and personal needs. But those needs will not be dispelled by reason and logic. Pierre Janet's history of psychological healing (*Psychological Healing*, Volumes 1 and 2 (translated by E. and C. Paul), Allen

and Unwin, 1925) demonstrates clearly that such denunciations may serve to modify, but will not eradicate, the responses to the continuous and widespread call for hope, support and consolation on the part of numerous people who are ready and willing to believe in what is offered for the purpose. Stripped of its scientific pretensions, psychoanalysis is merely an elaborate example of the many forms of psychotherapeutic activity that arise in response to these demands. The counsellors, the scientists, the cultists of various persuasions and some 'alternative' medical practitioners all testify to this phenomenon. As medicine attempts to shake off the trappings of magic and theology it is having to accept the old adage that while the physician's scepticism may be the best means of attaining truth, the patient's faith is often needed to attain health. □

Michael Shepherd is at the Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.

In for a penny, in for a pound

Norman Myers

Linking the Natural Environment and the Economy. Edited by C. Folke and T. Kaberger. Kluwer. 1991. Pp. 305. Dfl. 150, \$100, £52.

Values for the Environment. By J. T. Winpenny. HMSO. 1991. Pp. 277. £14.95.

ENVIRONMENTALISTS have much to learn from economists. For instance, can a natural resource such as a species really be 'beyond value', as many conservationists claim? Americans gladly pay \$15 million to save the California condor, but they might well jib at \$15 billion. Similarly, that key tool of economics, marginal analysis, has much to say about value where it most counts, namely at the margin: we would presumably pay more to save the tenth last panda than the thousandth last.

Conversely, economists have much to learn from environmentalists. Economics has little to say about the long-term future; a discount rate of 10 per cent effectively proclaims there is no future beyond seven years. Yet some of today's environmental assaults will leave their effects for centuries and millennia — in the case of mass extinction of species, for millions of years. Moreover, many economists assume that the welfare of future generations will continue to be better than that of the previous generation, so we should not worry unduly about the problems we may be passing

on: our descendants will find a way to sort them out, just as their forebears have always done. This thesis no longer holds when present-day environmental degradation looks set to deplete the world of many of its crucial natural resources. Most importantly, economics is largely based on an incrementalist approach to analysis, where situations change little by little rather than by big jumps. But ecosystems often undergo large shifts in make-up and dynamics, as is likely to be the case today as humans irreversibly transform the planetary ecosystem itself.

In short, it seems that economists and environmentalists do not talk enough to each other. They have been urged to compare notes for decades, and on all kinds of cogent grounds. But in my experience, the gulf of misconception is greater than ever.

All the more welcome, then, are these two books, each of which tries to bridge the gap. *Linking the Natural Environment and the Economy*, edited by two Swedish aficionados of the fields, is the more erudite and professional, dealing with "the dependence of socioeconomic systems on healthy environments and functional ecosystems". It seeks to demonstrate that "it is no longer possible to take environmental goods and services for granted". Instead they must be methodically taken into account by decision-makers at all levels. The book does a sound job, with much fine-grain analysis, while eschewing econometric minutiae.

The first part delineates several types of ecological and economic linkages, assessing their scope and importance through factors such as institutions, en-

vironmental effects and energy flows. The second part presents empirical analyses of the part played by environmental resources in economics, and contains examples from landscape change, agriculture, wetlands, and coastal and marine systems. Useful measurements such as plant growth, industrial energy, cropland nutrients, grain yields, soil erosion, eutrophication of water bodies and outputs of fisheries are invoked throughout, and there is a preliminary evaluation of trade-offs between competitive forms of land use. The third part looks at human impacts on environments in developing countries, through assessment of, for example, pesticidal pollution in Kenya and conservation in Central America.

The final chapter provides an overview of the eight case studies and their findings. This is the most illuminating chapter because it attempts, with some success, to define operationally the key concept of environmental economics, sustainable development.

By contrast, Winpenny adopts a sternly pragmatic approach. The author is a development economist with extensive field experience, and he hopes that his book will be "of operational value in project appraisal and, ultimately, policy formulation"; his intended reader is "an economist working with project responsibilities". He quickly plunges into a summary consideration of important habitats (such as watersheds and rainforests, as well as urban and industrial ones), and follows this with a review of evaluation techniques including cost-benefit analysis and "environmental benefit estimators". In the second half of the book he deals with practical applications of economics to, for example, desertification, loss of biodiversity, pollution and climate change, and includes sections on project appraisal and policy adjustment. All this is supported by many examples and small case studies.

The book is a fine analytical aid for project managers. But Winpenny seems to accept the situation 'outside the window' as given, with virtually no critical assessment of whether we are right about general environmental trends, or of what environmental economics as a discipline can offer if we decide we are not.

Both books can be commended to those trying to straddle the unfortunate divide between economics and environment; they are especially timely in the light of the United Nations Conference on Environment and Development to be held in June 1992, where interdisciplinary and intersectoral linkages will be central to the deliberations of the world's nations. □

Norman Myers is a consultant in environment and development at Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK.

Prophet of our times

John M. Charap

Sakharov Remembered: A Tribute by Friends and Colleagues. Edited by Sidney D. Drell and Sergei P. Kapitza. *Hilger/American Institute of Physics*: 1991. Pp. 303. £19.95, \$29.95 (hbk); £13.50, \$19.95 (pbk).

Andrei Sakharov: Facets of a Life. Editions Frontières: 1991. Pp. 730. FF295, \$50, £27.

JUST hours before he died in December 1989, Andrei Sakharov had addressed the Congress of People's Deputies, pleading once again the urgent need for reform, political pluralism and constitutional change. Three years earlier, Mikhail Gorbachev had telephoned to tell him "resume your patriotic work", thus ending his seven years of exile in Gorky. Sakharov had exhausted himself arguing for radical *perestroika*. His 'draft constitution' was never adopted, although it included reforms that are now being introduced as a result of the convulsions that have transformed the Soviet Union since his death.

The essays in *Sakharov Remembered* (many of them reprinted from special issues of the American Institute of Physics journal *Physics Today* and the Soviet Academy of Sciences magazine, *Priroda*) complement Sakharov's own *Memoirs* (Hutchinson, 1990) and *Moscow and Beyond* (Knopf, 1991). Susan Eisenhower and Roald Sagdeev believe that the Sakharov phenomenon (a phrase first used by his teacher, Igor Tamm) is virtually certain to attract the interests of generations of scholars, historians and biographers. These essays show why.

Sakharov's name was unknown in the West before his election to the Soviet Academy of Sciences at the unprecedentedly young age of 32. Even in the Soviet Union his work was known only to those with access to the secrets of Sarovskaya Pustyn (which Sakharov himself, with scrupulous respect for state security, always called simply "the installation"). He had "worked with enthusiasm" in this "Soviet Los Alamos" since 1948, earning the soubriquet 'father of the Soviet H-bomb'; for this he was elected to the Soviet Academy of Sciences, and honours were heaped on him by the State. Yu A. Romanov and V. I. Ritus, his colleagues at the academy, give what I believe to be the first declassified description of Sakharov's H-bomb design — a spherical, layered structure of ${}^6\text{LiD}$ and ${}^{235}\text{U}$ (or ${}^{239}\text{Pu}$) surrounded with unenriched natu-

ral uranium (${}^6\text{Li}$ generates tritium under bombardment from fast neutrons produced in the ${}^{235}\text{U}$ fission reaction, and this in turn fuses with the deuterium). This 'layer cake' leads to what the authors call "sakharization", enhanced yield through compression of the light elements; the natural uranium cladding contributes to this because of its high density, but also further enhances the energy released by fast-neutron-induced fission. Hans Bethe confirms that it was this "third idea" of Sakharov which in 1955 brought the Soviet Union level with the United States in H-bomb technology.

The announcement of the first Soviet H-bomb test coincided with Sakharov's election to the Soviet Academy of Sciences, and an acute observer might have



Sakharov with his wife, Klavidia Alekseyevna Vikhireva, and their children in the 1960s.

seen the connection. But as I. N. Golovin and V. D. Shafranov recall, it was a seminar on controlled thermonuclear fusion given at Harwell by Igor Kurchatov in 1956 that brought Sakharov's name to the attention of a wider international community. In his talk, Kurchatov described the tokamak design for magnetic confinement, which was conceived and elaborated by Sakharov and Tamm. As the magnetic thermonuclear reactor, this design has had recent success in the approach to controlled fusion at the Joint European Torus (JET) in the United Kingdom.

Sakharov worked on a whole range of problems related to fusion: muon catalysis (on which he wrote as early as 1948); generation of superstrong (hundreds of tesla) pulsed magnetic fields by "magnetic cumulation", converting the chemical energy of an explosive into magnetic energy by compressing a solenoid; the proposed initiation of fusion in deuterium using a pulsed laser. But at the same time he worked in fields far removed from applied physics. In 1965, he speculated that quantum fluctuations might be the seeding inhomogeneities responsible for the formation of galaxies, and two years later in his paper on induced gravity he showed how Einstein's theory might be an effect of vacuum fluctuations in quantum fields on a background metric. Perhaps his most remarkable work was on the baryon asymmetry of

the Universe, in which he suggested that the proton might be unstable: non-conservation of baryon number, together with C and CP violation could, during an epoch of thermal disequilibrium, then generate an excess of baryons even from symmetric initial conditions. His work on particle physics (for example, on meson masses) and cosmology continued up to, and indeed into, the time of his exile.

And of course, what makes the Sakharov phenomenon of such compelling interest is not his scientific achievement, but the events that led to his exile; his extraordinary courage and endurance, which earned him admiration and respect far transcending what his scientific achievements alone would have brought; and the moral leadership that

he gave through his persistent advocacy of human rights, of an end to nuclear confrontation and of the cause of freedom. The essays in *Sakharov Remembered* trace the development of his increasingly bold and open commitment to these causes, which would inevitably bring him into direct confrontation with the state and its masters. Maybe his childhood and upbringing prepared the ground. But it was in the installation, where growing anguish over the deaths caused by fallout from atmospheric nuclear tests brought him into direct conflict with Nikita Khrushchev, that he first set out on the path that would lead to his Nobel Peace Prize citation as "the spokesman for the conscience of mankind".

Sakharov's efforts to halt the tests met at first with crude rejection: but he persisted, with arguments backed by calculations of the biological and genetic damage that continued tests would cause. He took these arguments to the very top by writing to Khrushchev, and was at last instrumental in securing the Limited Test Ban Treaty of 1963. His conviction that the threat of nuclear holocaust was of overriding importance never left him. "The elimination of that threat takes unquestionable priority over all other problems in international relations", as Sidney Drell and Lev Okun quote him in their excellent essay on "Physics, the bomb and human rights".

Reproduced from Sakharov Remembered.

That threat may have receded, but we still live under its shadow.

Sakharov wrote again to Khrushchev in 1964, this time on an issue for which he had been slanderously attacked in *Izvestiya*, and criticized at the plenum of the Central Committee of the Communist Party of the Soviet union (CCPSU). The issue? His crucial speech that blocked the election of N. I. Nuzhdin to the Soviet Academy of Sciences. For Nuzhdin was a "faithful companion" of Trofim Lysenko, who regarded Sakharov's intervention as "a criminal matter". And it is reported that Khrushchev was so furious that he threatened to abolish the academy. (This did not occur; Khrushchev was dismissed from office in October 1964.) The pernicious sway of lysenkoism over Soviet genetics was broken by Sakharov's speech. B. M. Bolotovskii extensively quotes eyewitness accounts given to him by Tamm and Sergei Syrovatskii, as well as Sakharov's own reply to *Izvestiya* and his letter to Khrushchev. (Bolotovskii was the Communist Party organizer in the theoretical department of the academy's Lebedev Institute (FIAN), assigned by the party committee to "oppose" Sakharov during his years there.)

Khrushchev's denunciation of Stalin at the Twentieth Congress of the CPSU, and Roy Medvedev's polemic *Let History Judge*, impressed Sakharov deeply. His strong humanitarian convictions now led him to increasingly public protest and action. In 1968, his *Reflections on Progress, Peaceful Coexistence and Intellectual Freedom* was distributed through *samizdat*, and as *tamizdat* some 20 million copies were printed abroad. This cost him his job at the installation, and although he was allowed to join FIAN, he was now in official disgrace.

In the course of his human-rights campaigns, he met and married Elena Bonner, and when he was refused permission to travel to Oslo it was she who received the 1975 Nobel Peace Prize on his behalf. The abuse and vilification in the Soviet press rose to a peak. For protesting against the invasion of Afghanistan, Leonid Brezhnev exiled Sakharov to Gorky by a decree that had no status under Soviet law. During his seven years there, he suffered the appalling indignities of force-feeding when he went on hunger strike, and the spite of the KGB.

Both these books recall the disgust and anger of the scientific community, the moral support of his colleagues at FIAN and the unquenchable strength of his personal concerns, which always moved swiftly from general principles to concrete particulars. But it is to *Facets of a Life* that one should turn for passion and immediacy. It is easy to find fault with this "physicists' tribute" assembled by the theoretical department at FIAN.

It is untidy and uneven, with no evidence of editorial organization apart from alphabetical ordering by author; the English is distinctly Slavic; the subject-matter is often repetitive; there is no index, making cross-reference between the 60 contributions difficult. But for all that, I cannot recommend it too highly. Here is Sakharov as seen by those who grew up with him and worked with him; those whose lives were intertwined with his, who took risks on his behalf or who feel guilty that they did not risk enough. Not an icon of a saint, but sketches towards a portrait of the man.

The clamorous, argumentative essays jostle for attention. Many of them read like evidence at a tribunal, and most are concerned with setting the record straight. For example, there are agonizing attempts to clarify how Sakharov passed certain papers from Gorky to Moscow, one of which was a letter to Anatoly Aleksandrov, president of the academy. This letter, written *in extremis* to appeal for help before he began a hunger strike to secure permission for Elena Bonner to travel abroad (to visit her family and for medical treatment), is a heart-rending document, reproduced in the essay by Vitaly Ginzburg. But it was delivered a year late because of the scruples and hesitation of the "couriers". Not only Ginzburg, but also Boris Altshuler, Dmitry Chernavsky and

Evgeny Feinberg give their versions of what led Sakharov to write: "I have learned some details of what was going on in Moscow during my hunger strike and understood (but not reconciled myself to) the reason of [sic] disappearance of one of my documents."

Boris Altshuler's article, aptly headed with a quotation from Andrei Sinyavsky's — "I shall speak straight because life is short" — is packed with examples of Sakharov's knowhow; the achievement of strategic objectives by concentrating on concrete particulars. Altshuler had himself been fired as a physicist for his human-rights activities, and worked for many years as a caretaker. Feinberg addresses "the future historian" who will "understand Andrei Dmitrievich better than we, his contemporaries, obsessed with the intertwined and sometimes contradictory emotions and opinions." He emphasizes that Sakharov's political views evolved; it was this that perhaps gave them greater strength and maturity. Many of the authors speak of Sakharov's total integrity, the simplicity of his lifestyle, his apartness; he was regarded by some as a prophet. He surely touched their lives, and the shaping and direction of *perestroika*. And so our lives too. □

John M. Charap is in the Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

Alien life

Frank J. Tipler

The Cosmic Water Hole. By Emmanuel Davoust. MIT Press: 1991. Pp. 206. \$19.95, £17.95.

NEXT year, on Columbus Day, US researchers will begin a ten-year, \$100-million programme to search for extraterrestrial intelligence. Davoust, a French astronomer, provides a very readable popular description of this and other efforts to detect radio transmissions made by extraterrestrial intelligence. But his book marks an important change from the vast numbers previously published in this genre: he admits that the search will probably fail because such intelligence probably does not exist. It is refreshing to see an astronomer join the biologists, who, according to the distinguished biologist Ernst Mayr, are "... almost uniformly skeptical of the probability of extraterrestrial intelligence".

Davoust accurately expresses the force of the Fermi paradox: if extraterrestrial intelligences exist, why are they not already here? Even our primitive rockets can reach the stars in 10^5 years. Computer experts (such as Hans Moravec in

Mind Children, Harvard University Press, 1988) tell us that within 50 years we should be able to make an intelligent computer capable of self-reproduction. The combination of our primitive rockets with a payload of self-reproducing robots would be sufficient to explore and colonize the Galaxy. With such a probe, this would take only a few hundred million years to accomplish. We know that if our own evolution is typical, then most civilizations in our Galaxy must have arisen billions of years ago. Thus most experts now admit that civilizations so advanced would be here if they wished to come.

Davoust devotes an entire chapter to possible sociological mechanisms that might prevent interstellar travel. But I think he neglects the crucial point: virtually any motivation we can imagine that would lead extraterrestrials to engage in interstellar radio communication with us would also motivate them to engage in interstellar travel. The above-mentioned radio search is for civilizations wanting to contact us, for our equipment is not powerful enough to eavesdrop. Although we do not know the motivations of all advanced civilizations, we thus do know the motivation of the civilizations we search for. But robot probes would achieve the aims of these civilizations much better than radio sig-

nals. For example, probes can contact civilizations that are not listening, that is, those that do not have radio technology. Probes can be used to explore and colonize uninhabited systems.

Carl Sagan has argued that "perhaps [extraterrestrial intelligences] just don't care to strip-mine every site in the Galaxy". In his books, Sagan himself shows that all communication has costs; as he repeatedly says, the bare fact of a

"Those engaging in radio searches like to argue that absence of evidence is not evidence of absence. I totally agree, but we have evidence: extraterrestrial intelligences are not here."

received signal from an alien civilization would change ours drastically. But the purpose of any communication is to change the knowledge of the person to whom the message is directed: to colonize a mind with memes (complexes of ideas). There is no fundamental distinction between colonization with memes and colonization with genes. But in the case of genes, it is at least possible to limit colonization to uninhabited systems. Meme colonization necessarily occurs in inhabited systems, and necessarily extinguishes other memes. It is necessarily imperialistic.

Davoust discusses at length the anthropologist Ben Finney's comparison of interstellar colonization with historical human migrations. Finney claims: "No specific migration has ever gone unchecked. Ecological barriers, the slowing or cessation of innovation, flagging motivation, or the opposition of those in the way of expansion have . . . stopped every . . . colonization movement so far." Finney infers that interstellar colonization would stop short of the entire Galaxy.

But Finney's own data indicate the opposite. The analogue of the ecological-innovation barrier is the lack of a suitable robot probe, and our own civilization is near to overcoming this. With a probe, there is no natural barrier to stop a colonizing species short of the entire Galaxy. By definition, there is no opposition of those in the way for the first intelligent species to evolve. Finney's data indicate that motivation flagged once the other three barriers made further expansion difficult. Finney's picture of the evolution of Polynesian society is exactly what John Barrow and I predicted (*The Anthropic Cosmological Principle*, Oxford University Press, 1986) would be the behaviour of a colonizing extraterrestrial intelligence: an *r*-strategy characterized by rapid expan-

sion in numbers would be typical of those at the frontier, whereas a *K*-strategy characterized by fluctuations in numbers around an equilibrium would be typical of those in the interior.

Those engaging in radio searches like to argue that absence of evidence is not evidence of absence. (Davoust repeats this slogan at least twice.) I totally agree, but we have evidence: extraterrestrial intelligences are not here. We just have to interpret this fact. Most astronomers cling to a belief in extraterrestrial intelligence against the evidence because of a philosophical principle: the copernican idea that our place in the cosmos must be completely typical. But we know this idea is false. The Universe is evolving: the cosmic radiation shows that there was once a time when no life existed because it was too hot. Thus, our place is atypical in time. In particular there must be a first civilization, and it happens to be ours.

Davoust does not mention Brandon Carter's argument, based on the weak anthropic principle, for the nonexistence of extraterrestrial intelligence. This is unfortunate, because some of the most interesting new developments in particle physics use Carter's argument, which is derived from the fact that the time it took to evolve intelligence on Earth is within a factor of two of the lifetime of the Sun. Carter explains this approximate equality by assuming that the average time needed to evolve intelligence on an Earth-like planet is actually much longer than the lifetime of Sunlike stars. Biological evolution will cease when the star of an Earth-like planet dies, because the dying star destroys its planet. But the longer evolution can proceed, the more likely it is that intelligence will evolve. Thus the most probable time for the appearance of intelligence would be near the end of the time that evolution has had to operate on an Earth-like planet; that is, we expect approximate equality between the Sun's lifetime and the time needed to evolve intelligence. By making Carter's argument quantitative, S. Weinberg obtained an upper bound to the cosmological constant, whereas M. Shaposhnikov (*Modern Physics Letters* 59, 2607; 1987) made a prediction of the Higgs boson mass. This prediction failed, but it is a fascinating thought that there may be a connection between the Higgs mass and the rarity of intelligent life in the Universe. Pursuing this idea would be far more scientifically productive than doomed-to-fail radio searches. The original French title of Davoust's book was *Silence au point d'eau*. Silence there will be. □

Frank J. Tipler is in the Department of Physics, Tulane University, New Orleans, Louisiana 70118, USA.

Mosquitoes and backbiting

Len Goodwin

The History of Yellow Fever: An Essay on the Birth of Tropical Medicine. By François Delaporte. MIT Press: 1991. Pp. 181. \$30.25, £20.25.

RESEARCH scientists are human, and like to receive credit for their discoveries. But when recalling the development of their hypotheses, they can forget or omit to mention whose ideas really triggered them off. François Delaporte has made a detailed epistemological study of the classic discovery of the mode of transmission of yellow fever, accounts of which differ between South and North America. Cuban historians affirm that Carlos Finlay deserves most of the credit, whereas in the United States Walter Reed and his team are the true heroes. Both, of course, made vital contributions, but Delaporte, by noting what they read and wrote, reveals some facts that lead him to give credit for the seminal ideas to four Britons — Patrick Manson, Ronald Ross, Herbert E. Durham and Walter Myers.

Finlay, who had Cuban, French and Scottish blood, was an ingenious observer who had advanced several theories for the transmission of yellow fever, such as that involving increased ammonia in the atmosphere. In 1880 he had access to two important North American documents, the 'Plymouth' and the Chaillé reports, which showed that a ship needed contact with a port for an outbreak of yellow fever to occur and that some time elapsed between exposure at a port and an outbreak on board — the germ was "not like the poison of smallpox, but is produced and developed outside the body". He was also aware through the British medical press of Manson's work in China, showing that microscopic larvae of filarial worms in a patient's blood taken up by a mosquito developed in the insect to an infective stage. Finlay made the important observation that the 'Culex mosquito' (*Aedes aegypti*) could bite 12 times and lay three batches of eggs before it died; it could therefore convey infective material from one person to another, leading him to put forward his theory of mosquito transmission of yellow fever. But numerous attempts at transference from a patient to human volunteers failed because he kept the mosquitoes for only 3–5 days between bites. Finlay was regarded as a crank and it was 20 years before he was vindicated.

By 1900, yellow fever had visited 100 US cities and had killed thousands of

soldiers in the Spanish–American War. Reed's team was sent to Cuba in June; their main concern was with the bacterium *Bacillus icteroides*, claimed by the Italian bacteriologist Giuseppe Sanarelli to be the cause of yellow fever, and they spent a month disproving this hypothesis. But in mid-July they were visited by Durham and Myers from the Liverpool School of Tropical Medicine. These workers were familiar with the recent

that the infective agent passed through a bacterial filter and took 12 days to reach the infective stage in the mosquito.

But then politics took over and the controversy that developed, ostensibly a dispute over scientific credit, actually reflected a struggle for power. The defeat of Spain had not brought freedom to Cuba; the Platt amendment, which placed the country under the safeguard of the United States, humiliated the

Room at the top

John W. Galloway

The Outer Cycle: Women in the Scientific Community. Edited by Harriet Zuckerman, Jonathan R. Cole and John T. Bruer. Norton: 1991. Pp. 350. \$24.95.

WOMEN feature much less prominently in science than men. This book, a collection of research papers and reviews, sets out to tell us why.

Of course, this lack of prominence is hardly peculiar to science — it is broadly true of most public and creative walks of life. But science at its best should surely subvert the established social order and lean towards enfranchisement. To find it apparently so reactionary in this respect is a bit disturbing. The phenomenon is a ready subject for sociological analysis, an opportunity that has here been enthusiastically embraced.

What exactly is the phenomenon that the authors explore and try to explain? One of their concerns is the consistent paucity of women in science's upper reaches. Women form only two to three per cent of the membership of any number of national academies of science (although in the United States the proportion is now rising and has reached a heady four-and-half per cent). And so it is perhaps not unexpected that the proportion of women Nobel science prize winners is about the same. The figure rises a little if you throw in the prizes for literature and peace. Further, the deficit cuts across social barriers. For example, in the United Kingdom there are few women Fellows of the Royal Society or women Judges. Yet Fellows represent a complete spectrum of social origins, whereas judges are drawn mainly from the better-off classes. So in this regard at least, women do seem to be sisters under their skins.

The other concern of the authors is that this low showing is not simply a result of not enough women entering science to begin with. The other day I visited a large Scottish cancer-research laboratory. Women clearly outnumbered men among the research staff. But the women tended to be young and the older staff tended to be men. In general, and this is largely the issue that the book addresses, women as a whole do progressively less well the higher up the scientific hierarchy you look — they publish less, achieve less and get paid less than their male counterparts. Similarly, in the United Kingdom, those educated at independent schools do better the further up they are in the legal hierarchy.



Second World War US Army poster warning of bilharzia — another debilitating tropical disease caused by parasitic worms. From *Bilharzia: A History of Imperial Tropical Medicine* by J. Farley. Published by Cambridge University Press, £40, \$59.50.

discoveries of Charles Laveran and Ross that the malaria parasite used the mosquito as an intermediate host, and took time to complete its cycle in the insect. They also knew from W. S. Carter in New Orleans that there was an interval of 2–3 weeks between primary and secondary cases of yellow fever, and in a report published in the *British Medical Journal* in 1900 they wrote: "This curious and somewhat prolonged interval is suggestive of a development of the infective factor in or about some agent or matter in the domicile. . . . The suggestion propounded by Dr C. Finlay of Havana some twenty years ago, that the disease was spread by means of mosquitoes hardly appears fanciful in the light of recent discoveries in ague convection. . . ."

The US team took action immediately. On 1 August, Reed, James Carroll and Jesse Lazear went to Finlay and obtained the eggs of his *Culex* mosquito. Reed was recalled to Washington, and the first, not very well controlled trials with mosquitoes were made by Lazear. Carroll was bitten on 27 August by an insect that had been fed on a yellow-fever patient on the second day of illness and that had been kept for 12 days; he had a sharp attack of fever. The rest of the story is well known — Lazear was bitten accidentally and died, and Reed took charge and showed conclusively

Cuban people and opened their eyes to the meaning of US intervention — the United States had supplanted Spain. The names of Reed and Finlay became convenient symbols for a clash of values, and mythical, political versions of the discovery were spread through the media. The two official pictures painted to celebrate the discovery, by Estevan Valderrama in Cuba and by Dean Cornwell in the United States, depict not real events but fictions.

Delaporte's study provokes thought as to the real origins of successful ideas. The translation from the French is good and retains a Gallic flavour with few hiccups. But you can't help feeling sorry for Finlay. Eccentric, devious and a bit of a showman, he had the brilliant, original conception of the mosquito as a vector but had no way of seeing it as a second host. With a little more luck, or a little less care, he might have kept one of his mosquitoes for a few more days: medical science would have been advanced 20 years; Finlay would have the credit; a Cuban volunteer, not Lazear, would have been martyred; and there might never have been a Walter Reed Army Institute of Research in Washington DC. □

Len Goodwin (formerly with the Wellcome Trust) is at Shepperlands Farm, Park Lane, Finchampstead, Wokingham, Berkshire RG11 4QF, UK.

There is plenty of evidence here (as there is elsewhere) for the poor showing of women in the scientific community, such as that provided in a comprehensive review by Harriet Zuckerman. But explanations for why they have such a peripheral status are thinner on the ground. Indeed, one of the book's strengths is the number of possible explanations that it turns upside down simply with relatively trivial observations.

Families

To take an obvious example, women with families use more of their energy in looking after their families than do men with families — as a consequence, men have more energy to spend on their research. But women with families are no less productive as researchers than are women without. The book might have added that many (perhaps all) of the most successful women scientists have or have had families. Dorothy Hodgkin in *Who's Who* lists her hobby as 'children'.

Historically, attitudes of men scientists towards women scientists were blatantly discriminatory. The treatment of Marie Curie by the French Academy of Sciences is as striking an example as any. She would have been the first woman member, but was not elected, despite having been a Nobel prize winner, as well as the first woman to hold a doctorate in France and be a professor at Sorbonne. Dorothy Hodgkin, even though a Fellow of the Royal Society at the age of 37, was denied membership of the University of Oxford's men-only chemical club, the Alembic. In any case, it took the Royal Society about 250 years to elect its first woman members. Even the US National Academy of Sciences took 63

years. And several well-known physicists, such as J. J. Thompson, Louis de Broglie and R. A. Millikan, are on record for having found it impossible to take women seriously as scientists. Of course, there might have been the odd woman who had a special scientific gift, but then in making this concession there was also the implication that pigs might fly.

But today things are different, aren't they? The formal barriers to women achieving success in science have all gone. Or do they still exist in the form of an "invisible obstacle race", as Joan Mason recently described it in *Nature* (353, 205; 19 September 1991)? It is a situation reminiscent of the abolition of slavery or the opening up of higher education to the working classes. Things are better, but nothing like as good as they ought to be. Those apparently being given opportunities do not seem to be able to take advantage of them.

We are obviously in the midst of a historical process, and an evolutionary analogy might be apt. Does the system discriminate against women simply because they are rare? The book provides evidence that this is so: unlike their male counterparts, women scientists tend to cite other women's papers. Also, as is made clear, women prefer some sciences to others, being much more oriented towards biology, psychology and sociology than to the mathematical or physical sciences (most mouse geneticists are women).

It is probably significant, therefore, that in the Royal Society, where cognitive psychology claims only about four fellows, two of them are women. How heavily is sociology represented in national academies of science? Medicine was the first 'science' in which women

started to make substantial inroads, beginning in the 1840s in the United States and happening 30 years later in the United Kingdom, and academic medicine must be the science in which women are most strongly represented. The medical profession seems to have been less discriminatory or to have mended its ways well before other professions, probably because of pressure by women for equal opportunities and a demand by other women to be treated by female doctors. Surveys in the United States, reported in the book, indicate that women who want to study medicine have not been discriminated against for the past 60 years. If there are fewer women in medicine it is because fewer have wanted to be in the profession. Of course, that raises another question, why don't they?

Differences

The book has some interesting things to say but it has a serious weakness. It gives little impression of what the business of being a scientist really entails. And given that the book's central thesis is that for an answer one must look to the structure of science as a social organization rather than to biology, I would have liked a much better feel for that structure. This is tied up with the fact that different sciences really are very different from one another: although the way in which women fare is in many respects similar among different sciences, there is no escaping the very real differences. This point is little more than touched on.

Although I now feel better informed to tackle the question of why there are so few successful women in science, I do not really feel much the wiser. Without doubt the lack of women in the scientific community is a very real phenomenon

ADVERTISEMENT

THE ARROW OF TIME

PETER COVENEY & ROGER HIGHFIELD

THE QUEST TO SOLVE SCIENCE'S GREATEST MYSTERY

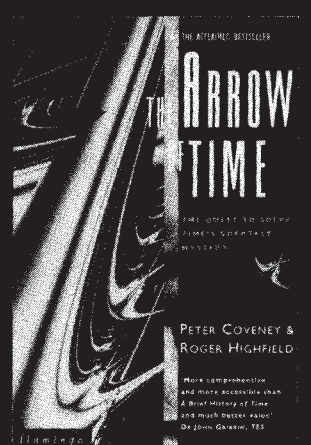
'More comprehensive and more accessible than
A Brief History of Time and much better value'

TIMES EDUCATIONAL SUPPLEMENT

'This is an important book...
I heartily commend this volume'

NEW SCIENTIST

NOW A BESTSELLING FLAMINGO PAPERBACK £5.99



and one that is unfortunately rather refractory to change. Indeed it is so concrete and robust that you feel there must be a readily and well-defined cause. Perversely, the book's message is that there isn't. □

John W. Galloway (formerly at the Cancer Research Campaign) is at the Nuffield Foundation, 28 Bedford Square, London WC1B 3EG, UK.

Rewriting the future

Walter Gratzer

Great Mambo Chicken and the Trans-human Condition: Science Slightly Over the Edge. By Ed Regis. Viking: 1991. Pp. 199. £16.99.

The Virgin and the Mousetrap: Essays in Search of the Soul of Science. By Chet Raymo. Viking: 1991. Pp. 308. \$18.95.

THE trouble with progress, as the philosopher said, is that it goes on too long. Ed Regis's excellent vade-mecum to the wilder shores of science gives evidence enough to chill your gizzard for the truth of this proposition. Here you will learn about plans for hijacking and populating comets as our world grows less friendly and more crowded, for moving the Earth into a more hospitable orbit by extending solar sails into space to gather thrust from sunlight and, since vaulting ambition may as well o'erleap itself as not, for transforming the system of sails into a nuclear fusion engine, powerful enough to move the Sun itself. A Harvard professor of physics develops a grand unified theory (or GUT) that he believes to fit the facts at last and finds to his ineffable dismay that it implies the instability of the proton; so in a space of 10^{31} years atomic matter will collapse like an underdone soufflé. Gloom and despair take possession of him — but never fear, for Freeman Dyson is at hand with a wheeze: all we need do is to generate a huge, cosmic energy flux (easily encompassed by turning a lot of matter into radiation) that will split open the closed Universe like a seed-pod, and then the collapse will be offset by endless expansion in another direction. (Dyson anyway does not hold with the idea of a closed Universe, which gives him, he says, a feeling of claustrophobia.)

Achieving immortality

Or we could tinker a little with our DNA and rewire ourselves for life at the near-absolute zero of outer space. We might even learn to hibernate, and then our periodic moments of consciousness

would seem, thanks to our arrested metabolism, like an eternity. James Thurber had a similar apprehension: "I'm sixty-three", he wrote to his publisher when he was sixty-three "and I guess that puts me in with the geriatrics, but if there were fifteen months in every year, I'd only be forty-three."

You may of course take the view that the future, when Earth is but a star that once had shone, can be left for your grandchildren to trouble themselves about, but, in California at least, the opinion seems to have taken hold that death is discretionary. There is, for instance, the alternative course of having your head detached and preserved in liquid nitrogen until such time as the trifling problems that still stand in the way of growing a disease-proof body from the neck down have been overcome. Cephalariums, full of bubbling vats of heads, already clutter the western seaboard. Some 25 years ago, as Regis reminds us, a paper appeared in this very journal with the good news that a cat brain, brought to room temperature after six months in the freezer, at once began to put out electrical signals little different from those generated by the living cat.

But stay: if you wait until you die — or rather deanimate, in the language favoured by the believers — before suffering your head to be frozen, then you may not be in your mental prime when the last trump sounds and you rise from the nitrogen. The jazz musician Eubie Blake remarked on his hundredth birthday that if he had known he was going to live so long he would have looked after himself better. Well, no need to lose sleep over such dilemmas because less conservative thinkers, such as Professor Moravac, will tell you that freezing heads is yesterday's technology. The reason, you ask? Why, because nanotechnology is around the corner, and 50 years from now, perhaps less, there will be molecular robots to surge through your brain, neuron by neuron, reading off stored electrical signals. A year or two's work will suffice to 'download' the contents of your brain onto disc, and then you may live for eternity, in a body with the form, materials and colours of your choice, experiencing by way of appropriate inputs all the sensations you desire, from the perpetual orgasm that is said to feature prominently on the Muslims' agenda for paradise, to irrigating your tonsils with bounteous tides of Chateau Yquem. Nor will your immortality be tried by earthquakes or nuclear explosions, for back-up copies of your disc — of you — will repose in safe places around (or beyond) the world.

If progress depends, as George Bernard Shaw suggested, on the unreason-

able man, you will find the whole *galère* in Regis's pages. Here is virtuoso science writing of rare vigour and mastery; and for those who enjoy being frightened it eclipses Mary Shelley and her Gothic nightmares. If, however, you find Regis's supercharged style too wearing in large doses, then perhaps Chet Raymo is your man. He is a professor of astronomy and columnist in the *Boston Globe*, and the essays gathered in his new volume are polished, civilized and literate discourses on a wide spectrum of scientific topics.

Selfish genes

Raymo is captivated by the night sky and enchanted with the *Escherichia coli* to which we are hosts — on so generous a scale indeed that your gut contains, as he calculates, as many as would stretch from Boston to San Francisco. He ranges from Aristarchus and the astronomy of the ancients to selfish genes and the anthropology of the Yanomano Indians of the Amazon, who slaughter each other with no compunction in the interests of sex; homicidal males are rewarded by more wives and many more progeny than pacifists. Are these, Raymo wonders, behavioural genes in action?

Raymo mourns the passing of Rube Goldberg, whose concerns, it seems, transcended those of our nearest British counterpart, Heath Robinson (the man who devised such aids to good living as a machine for the removal of gravy stains from gravel drives). For Goldberg warned against the "gadget-strewn paths of civilisation", among which Raymo numbers nuclear power stations and the monstrous products of the Star Wars programme. There are 21 essays in this collection, all of them cunningly crafted and a pleasure to read, even if here and there the social concern (not to say the PC usage) is laid on a touch too liberally. Raymo is drawn to the philippics of Chargaff, whose apocalyptic visions of a society in thrall to the new genetics find little resonance in the imaginations of other biologists. And you may also stub your toe on an occasional abomination such as "person year" (it followed for me a recent sighting in this country of a ploughperson's lunch); but perhaps one can blame such lapses on a malign subeditor. □

Walter Gratzer is in the MRC Muscle and Cell Motility Unit, King's College, Drury Lane, London WC2B 5RL, UK.

Spring Books

Nature's next review supplement will be Spring Books, which will appear in the 16 April 1992 issue.